The Dissertation Committee for Yihong Zhou
certifies that this is the approved version of the following dissertation:

Spatial Usage and Power Control in Multihop Wireless Networks

Committee:

Scott M. Nettles, Supervisor

Gustavo De Veciana

Robert W. Heath

Craig Chase

Lili Qiu

Spatial Usage and Power Control in Multihop Wireless Networks

by

Yihong Zhou, B.S; M.S

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2006

Spatial Usage and Power Control in Multihop Wireless Networks

Publication No. _____

Yihong Zhou, Ph.D
The University of Texas at Austin, 2006

Supervisor: Scott M. Nettles

In wireless networks, because of co-channel interference, concurrent transmitters must be chosen such that they provide a minimal amount of interference to each other. In a simple path-loss based propagation model, this implies that the concurrent transmitters and receivers must be spaced sufficiently far apart. When network traffic is high, space becomes a limited resource for which every node has to compete. Thus, improving spatial reuse among wireless nodes or reducing spatial usage of wireless transmissions is crucial to improving overall network throughput.

Among the techniques for improving spatial reuse, transmit power control is fundamental. In this dissertation, we first analyze the impact of transmit power on

potential network throughput. To do this, we propose a spatial usage metric and then investigate the impact of transmit power on the spatial usage of single and multihop communications.

Motivated by our analysis, we propose a *Media Access Control* (MAC) and physical layer power control scheme, *Optimized Transmit Power* (OTP), to balance the spatial usage of each individual transmission and co-channel interference. This scheme assumes the worst possible interference at the receiver and reduces transmit power to be just great enough to guarantee reliable signal reception. Further study shows that OTP is overly conservative, because the worst case interference does not occur much of the time. Therefore, we develop an *Enhanced OTP* (EOTP) to tradeoff a possible occasional collision for lower power and better spatial usage. Our simulation results show that EOTP outperforms OTP and both schemes improve overall network throughput to a moderate or significant degree.

Because MAC layer power control schemes favor short sender-receiver distances, we study a mini-hop routing strategy that discovers routes consisting of short distance hops and develop a *Mini-Hop Routing* (MHR) protocol. When combining MHR with EOTP, network performance, including throughput, end-to-end packet delivery latency, and routing overhead, is improved substantially.

Finally, we study a load-sensitive routing strategy that bypasses hot spots and utilizes idle space. Our investigation demonstrates that the existing blind flooding technique is able to circumvent hot spots to a significant degree. Load-sensitive routing outperforms the blind flooding technique substantially only when flow lifetime is short or node mobility is high.

Thesis Statement

By properly adjusting transmit power with the assistance of enhanced MAC and network layer protocols, we can improve spatial reuse in multihop wireless networks, resulting in improved overall network performance.

Contents

Chapter 1

Introduction

With the explosion of demand for wireless communication services, wireless networking has received significant attention over the last few years. Based on how packets are forwarded, wireless networks can be divided into two categories: single-hop and multihop wireless networks. In single-hop networks, after one wireless hop, packets are forwarded by a centralized usually wired infrastructure. Typical examples include cellular networks and wireless LANs based on the IEEE 802.11 *Point Coordination Function* (PCF) mode [1]. In multihop wireless networks, packets are forwarded by multiple wireless nodes. Typical examples include mobile ad hoc networks [2], sensor networks [3], and mesh networks [4], etc. At present, many aspects of multihop wireless networking have not been thoroughly investigated.

Unlike the wired channel, which has Giga-bits per second (Gbps) or even higher transmission capacity, the wireless channel capacity is restricted by limited spectrum, interference, and signal fading among other factors. Thus, wireless networks are challenged to support high-throughput applications, and leveraging various techniques to improve network performance and in particular throughput is crucial for extending their applicability. Many factors, such as physical layer techniques, transmit power level, the medium access mechanism, and the network layer

routing scheme, can affect network throughput, and their impact and interactions are not intuitive. Therefore, we seek a common factor, or unifier, that can be the bridge between these factors and network throughput. This common factor should have a straightforward relationship with these factors and, at the same time, be an indication of potential network throughput.

We claim that spatial usage (or spatial reuse), which measures the efficiency of network space utilization, is such a common factor. The intuition behind the relationship between spatial usage and potential network throughput is straightforward. When multiple wireless nodes share a common communication channel, concurrent transmissions interfere with each other. Received signals can be correctly decoded only when the ratio of the received power level to the level of the interference and noise is above a certain threshold. Hence, in a simple path-loss based propagation model, to guarantee reliable signal reception, concurrent transmitters must be kept a certain distance apart, and the number of concurrent transmissions is constrained per unit area. In particular, when the traffic load is high, space becomes a limited resource for which every node has to compete. Therefore, efficient spatial reuse means more potential concurrent transmissions and higher potential network throughput. Schemes that improve spatial reuse potentially improve overall network throughput.

Among the techniques for improving spatial reuse, transmit power control is a simple and fundamental one. Reducing transmit power level decreases co-channel interference and the "space" each transmission occupies. Therefore, concurrent transmission can be packed more tightly, which means better spatial reuse and higher network throughput. In addition to improving spatial reuse and network throughput, power control reduces channel contention and, therefore, end-to-end packet delivery latency. Although improving end-to-end packet delivery latency is not the major goal of our study, we use it as one of the metrics to evaluate our proposed protocols.

In a basic sense, power control relates to issues, such as radio propagation and signal reception, that are most closely coupled to the physical layer. But, power control has significant interactions with the *Medium Access Control* (MAC) layer. The MAC is part of the *data link control* (DLC) layer and comprises all mechanisms that regulate users access to the medium. From the perspective of spatial usage, the MAC allocates limited network space resources among multiple nodes. Power control and MAC mechanisms interact with each other such that, on one hand, transmit power determines link quality and the schedule of concurrent transmissions; and, on the other hand, enhanced MAC protocols can help to optimize transmit power for each individual transmission.

Power control plays a role at the network layer as well. The basic functionality of the network layer is to discover and maintain routes between sources and destinations. The transmit power determines the set of possible neighbors a node can reach and hence the potential set of next hops and the overall network connectivity. On the other hand, sophisticated routing techniques can help power control schemes improve network performance. For example, shortest-path routing chooses routes that contain the least number of hops each of which spans a long distance. Under this routing technique, power control cannot gain much of an advantage, because the chances for reducing transmit power are restricted. Instead, a routing technique that chooses routes containing shorter hops may provide more opportunities to reduce transmit power level substantially and improve spatial reuse.

In summary, this thesis focuses on improving spatial reuse by developing power control schemes at the MAC and network layer. In particular, we focus our studies on multihop wireless networks, although single-hop wireless networks form a special case of our more general studies.

1.1 Thesis Statement

My thesis is:

By properly adjusting transmit power with the assistance of enhanced MAC and network layer protocols, we can improve spatial reuse in multihop wireless networks, resulting in improved overall network performance.

1.2 Goals and Approaches

Motivated by improving spatial reuse and network performance with transmit power control, this thesis has two goals:

1. To understand spatial usage, its relation to network throughput, and how it varies as a function of transmit power.

2. To use this understanding to design enhanced MAC and network layer protocols that improve spatial reuse and network performance.

There are three steps to achieve our goals.

First, we define a spatial usage metric, Ω, and study its relationship to potential network throughput. By bounding each term of the metric as a function of the transmit power, we find the upperbound and lowerbound of Ω as a function of transmit power. Further, employing the relationship between Ω and the potential network throughput, the impact of the transmit power on the potential network throughput is revealed. To validate our theoretical analysis, we perform experiments using the *Network Simulator* (ns2) [5, 6]. Our initial experimental results do not match our theoretical predictions. The reason for this mismatch is the border effect that often occurs when dealing with spatial data. The border effect is caused by the fact that the behavior and characteristics of nodes on the border of the network are different than those inside. To eliminate the border effect, we

propose two approaches: enlarging the network scale, and using toriodial geometry. Applying these two approaches to our experiments, we obtain a close match between the experimental results and our theoretical predictions.

Second, taking advantage of the insights gained by our analysis, we focus on research that restricts the strategies for improving spatial reuse to the MAC layer. We propose two MAC layer power control schemes, *Optimized Transmit Power* (OTP) and *Enhanced Optimized Transmit Power* (EOTP), both of which adjust transmit power for each individual transmission. OTP always assumes the worst possible interference at the receiver and tunes the transmit power such that it is just great enough to guarantee reliable signal reception. Simulations show that OTP improves overall network throughput to a moderate or even significant degree. Further study shows that OTP is overly conservative, because the worst interference case does not occur all the time. Therefore, we propose EOTP to adjust transmit power by tracking the actual interference level at the receiver. In doing this, we tradeoff a possible occasional collision with lower power. Simulations show that EOTP may improve overall network performance substantially.

Third, we study new routing protocols to assist power control schemes in improving spatial reuse and network performance. Taking advantage of the insights gained by our analysis of spatial reuse, we propose the Mini-Hop Routing (MHR) protocol, which attempts to discover routes consisting of many short hops. Combining MHR with appropriate MAC power control scheme, spatial reuse and network performance are further improved compared to the MAC only approach. Furthermore, to fully utilize the networks spatial resource, we investigate load-sensitive routing, which discovers routes that bypass hot spots and utilize idle space. Our study shows that the blind flooding technique, which is employed by most on-demand wireless routing protocols, is load-sensitive to a certain degree. To reinforce the load-sensitivity, we propose two routing schemes, Delaying Propagating RREQ

(DPR) and Tagged RREQ (TGR) routings. Our simulations show that both DPR and TGR outperform the blind-flooding technique substantially only when traffic flow lifetime is relatively short or node mobility is relatively high. This is because non-congested routes are stable while congested routes are likely to be broken due to intense channel contention, resulting in the blind flooding technique eventually converging routes to idle space.

A final point is that the work in this thesis is not fundamentally cross layer design. Even though a small amount information needs to be passed between the network, MAC and physical layer, we do not propose a significant merging of the layers.

1.3 System Assumptions

Our work is based on the following assumptions:

1. All wireless nodes are homogenous and use omni-directional antennas.

2. All wireless nodes share a common communication channel. Even if they use different channels, the spatial usage analysis in each sub-channel is identical.

3. We base much of our analysis on the random access mechanism, *carrier sense multiple access* (CSMA).

4. In terms of radio propagation, we only model large-scale path loss but ignore small-scale fading and shadowing effects. The large-scale path loss models are discussed in Chapter 2 and they are mainly used in Chapter 4.

5. We assume fixed modulation and coding schemes at senders, e.g., the senders do not adapt sending rate as transmit power varies.

6. We assume that the physical layer is sophisticated enough to measure the average power of a received signal and the average noise-and-interference level [7]. Further discussion of this issue is found in Chapter 2.

In subsequent Chapters, we discuss the impact on our work if the last two assumptions are relaxed and present potential future work to cope with these issues. Additionally, in Chapter 8, we discuss the impact of small-scale fading.

1.4 Scope of the Dissertation

Our analysis and proposed protocols are sensitive to a variety of network characteristics, such as traffic patterns, network topology, node placement and distribution, etc. In fact, it seems likely that any scheme seeking to optimize spatial usage will be sensitive to many of these factors. These factors depend on exactly how a wireless network is deployed and used. Unfortunately, at present, very few multihop wireless networks have been deployed and thus, very little is really known about these key network characteristics. Even when more networks are deployed, it is likely that different networks will have very different topologies and traffic patterns. For example, in a sensor network that monitors environmental temperature, nodes would be mostly static and evenly distributed, traffic flows short-lived, and each individual packet would be small. On the other hand, in a home-entertainment network whose purpose is to transfer audio/video content, traffic flows would be long-lived and each individual packet would be large. Finally at a conference, when people gather into groups, the network topology is clustered, traffic flows may demonstrate locality and mobility will perhaps be a factor.

Since we do not know the key network characteristics, designing a protocol that yields the "the best" performance is essentially impossible and, in fact, probably there is not a protocol that is optimal for all reasonable network scenarios. Our thesis is that power control can improve network performance. Our strategy for

demonstrating our thesis is to show that feasible and plausible protocols exist, that they do improve network performance by improving spatial reuse, and that they work on some small but plausible set of topologies and traffic patterns. Given our limited knowledge this seems only reasonable and so we do not attempt to explore the whole protocol design space and certainly not the whole set of network scenarios.

We will see a number of examples where key design parameters do depend on details of the network. For example, in Chapter 7, the load-sensitive routing protocol, TGR, uses a channel utilization threshold above which a node is considered congested. Inappropriate choice of this threshold would lead to a node being unable to distinguish non-congested status from congested status properly. To identify the optimal threshold that yields the highest network throughput, we did extensive simulations. We found that the optimal threshold varies with the degree of congestion and the nature of the flows. We further found that if flows are long lived and mobility low, existing, naive algorithms actually perform well.

Nevertheless, through spatial usage analysis, protocol design, and limited experimentation, we gain insights that can be used by designers as the nature of these networks become more clear. For example, we find that the performance of power control schemes are sensitive to network scale. We leave the full exploration of the design space as future work for when key network characteristics are more clear.

This dissertation focuses on improving network throughput using MAC layer power control schemes and advanced network layer routing strategies. Although we considered other performance metrics, such as end-to-end packet delivery latency and routing overhead, energy consumption is beyond the scope of our study.

We ignore the existence of fast fading and assume that the distance between a sender and receiver, d, can be derived given the transmit and receive power levels. This assumption occurs in the OTP and mini-hop routing protocol designs. It is

not crucial for three reasons. First, the power level proposed by OTP, P_{otp}, is eventually used mainly as an upperbound in the EOTP scheme. We will show that P_{otp} is overly conservative. Thus, a slight variation of P_{otp} caused by errors in the distance calculation just gives EOTP either a tighter or a looser upperbound. Second, for the mini-hop routing protocol, we propose an alternate solution of setting the link cost to be a function of the data packet transmit power instead of the sender-receiver distance. Third, the sender-receiver distance could be obtained using other existing techniques, such as GPS or more sophisticated physical layer schemes. Nevertheless, investigating the impact of fast fading on our protocols is an attractive future research topic.

A final point is that, in much of the dissertation, we assume a fixed transmission rate. Although there are some wireless networks that are strictly fixed rate, in particular some sensor networks, the primary motivation for this assumption is to factor the problem at hand into tractable pieces. In particular, our primary goal is to investigate the impact of transmit power on network performance. In Chapter 3, we relax this assumption and introduced the factor of rate into our spatial usage analysis. The preliminary analysis shows that when space is at a premium, it may be more desirable to reduce transmit power than to increase rate. Further, in each subsequent chapter, we discuss the impact of using rate adaptation on the protocol being investigated. In part because it is clearly dependent on the (currently unknown) topologies and traffic patterns of multihop wireless networks, we leave a thorough investigation on the combined effect of adaptive-power and adaptive-rate for future work.

1.5 Road Map

The rest of this thesis is organized as follows: We introduce some basic background knowledge and related work in Chapter 2. We analyze spatial usage as a function of

transmit power in Chapter 3. In Chapter 4, we propose a MAC layer power control scheme, OTP, to improve spatial reuse and network throughput. In Chapter 5, we enhance OTP and develop EOTP. In Chapter 6, we design a mini-hop routing protocol, MHR. In Chapter 7, we develop two load-sensitive routing protocols, DPR and TGR. Chapter 8 presents the potential future work and, finally, Chapter 9 concludes with our contributions.

In addition, Appendix A proves two features of shortest-path routing and Appendix B presents Monte Carlo simulations for the hop-distance distribution of shortest-path routing. In Appendix C, we derive the spatial reuse factor for a chain traffic flow. In Appendix D, we derive the spatial reuse factor for multiple parallel traffic flows.

Chapter 2

Background Knowledge

In this Chapter, we introduce some basic background knowledge that is needed for the rest of the thesis. Since our goal is to propose enhanced MAC power control schemes and routing protocols to improve spatial reuse, we introduce background knowledge in three major areas:

1. Radio propagation, signal reception and SINR estimation etc.

2. Basic MAC operations including the random medium access mechanism, an idealized three-range model of wireless transmission, the exposed and hidden node problems, and the IEEE 802.11 MAC protocol *Distributed Coordination Function* (DCF) mode [1].

3. Routing protocols for multihop wireless networks.

At the end of this Chapter, we also review related work. This review is a high level introduction of current research status on spatial reuse and power control. Specific related work is presented in more details in each individual chapter.

2.1 Radio Propagation Models

To understand how to control transmit power, we must first understand the basic knowledge about radio propagation. Because physical layer techniques are not the topic of this thesis, we abstract radio propagation to large-scale path loss and ignore shadowing effects as well as small-scale fading.

In large-scale path loss models, the average received power level, P_r, is a function of the Transmitter-Receiver (T-R) distance, d. That is

$$P_r = \beta \frac{P_t}{d^\alpha} \ , \tag{2.1}$$

where P_t is the transmit power, α is the path loss exponent, and β is a constant.

In particular, the free space propagation model and the two-ray propagation model are commonly used in wireless communication analysis. The free space propagation model is used when the transmitter and receiver have a clear, unobstructed line-of-sight path and there is no ground reflection between them. The path loss exponent α is 2, and the received power level is given by the Friis free space propagation model [8]

$$P_r = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L} \ , \tag{2.2}$$

where G_t and G_r are the antenna gain of transmitter and receiver respectively, λ is the wavelength in meters, and L is the system loss factor not related to propagation ($L \geq 1$). This is an ideal model that captures the basic idea of radio propagation. It is applicable in scenarios that the transmitter and receiver have a clear, unobstructed link-of-sight path and that there are no ground reflections, surrounding objects and obstacles. For example, satellite communications.

The two-ray propagation model considers both the direct path and a ground reflected propagation path between the transmitter and receiver [8]. The path loss

exponent α is 4, and the received power level is

$$P_r = \frac{P_t G_t G_r h_t^2 h_r^2}{d^4} \ , \tag{2.3}$$

where h_t is the height of transmitter antenna and h_r is the height of receiver antenna. This model is an approximation that holds for long transmitter-receiver distance. It considers two paths, the line-of-sight propagation and a ground reflection.

There are other large-scale propagation models corresponding to different environments and thus different path loss exponents. We use (2.2) and (2.3) in Chapter 4 to calculate data transmit power, P_{otp}. Using a different model would not affect the protocol desgin in a fundamental way, since the basic idea of OTP is not affected by the specific constants or path loss exponents. In fact, we could have obtained the same results using the general path-loss model (2.1), which would make our derivation even simpler. We used the other two models because they are consistent with the NS2 simulator [5, 6].

2.2 Signal Reception

It is also important to understand what happens when a radio attempts to receive a packet. In reality, a receiver simply attempts to decode the detected signals. As long as the checksum is valid, the signal is said to be correctly decoded. In theoretical analysis, we use a model that accounts for when receptions are likely to succeed. Signals are correctly decoded as long as the following two conditions are satisfied. First, the received power level must be greater than a certain threshold, $\mathcal{T}_{rx}$. Second, the ratio of the received power level and the level of interference and noise, i.e., the *signal to interference and noise ratio* (SINR), at the receiver must be greater than a certain threshold, $SINR_0$. This is referred to as the capture effect. When another transmission drops the SINR below this threshold, we say a collision occurs.

2.3 Carrier Sense Multiple Access

All of this dissertation is based on MAC protocols that use *carrier sense multiple access* (CSMA) as part of their basic mechanism. CSMA has a long history, with roots dating to long before the advent of packet radio [9]. It is a kind of *Time Division Multiplexing* (TDM). The basic idea is to listen for a carrier before accessing the medium, and then access the medium only if the channel is idle. CSMA is usually combined with a back-off scheme in the case of a busy medium so that some fairness can be achieved among competing nodes.

Based on what to do when the medium is busy, several versions of CSMA exist: non-persistent CSMA, p-persistent CSMA, and 1-persistent CSMA [10]. Additionally, combined with *Collision Detection* (CD) and *Collision Avoidance* (CA), CSMA results in other schemes: CSMA/CD and CSMA/CA [10, 11]. Both collision detection and collision avoidance are used to improve the performance of CSMA. Collision detection terminates transmission as soon as a collision is detected, while collision avoidance attempts to avoid collisions by reserving the network for individual transmitters. Collision avoidance is used in media such as radio, where reliable collision detection is not possible. Nowadays, CSMA/CA is widely used in WLANs.

In theoretical analysis, we use a model that accounts for when a carrier is likely to be detected. That is, a carrier can be detected as long as the received power level is greater than the carrier sense threshold, $\mathcal{T}_{cs}$.

2.4 Three-Range Model

In order to facilitate our theoretical analysis, we use an idealized three-range model [12, 13] to describe CSMA based wireless transmissions. In this model, three thresholds are important: the carrier sensing threshold, $\mathcal{T}_{cs}$, the receiving threshold, $\mathcal{T}_{rx}$, and SINR threshold, $SINR_0$. All of these are characteristics of the wireless nodes phys-

ical layer implementation.

Closely related to these thresholds are three ranges. When a transmitter X transmits to a receiver Y, these three ranges are crucial: the *carrier sense range* (CSR) around X, the *transmit range* (TR) around X, and the *interference range* (IR) around Y. We explain each of these ranges in turn below.

2.4.1 The Carrier Sense Range

The carrier sense range of the transmitter X, CSR_x, is the area around X, within which a node can detect the transmitted signal. That is, the signal power level P_r at the node is greater than $\mathcal{T}_{cs}$. In a CSMA MAC, all nodes within the CSR of an ongoing transmission defer their own transmissions until the channel is idle. According to this definition and the large scale propagation models discussed earlier [8], the radius of the CSR is

$$R_{cs} \propto \sqrt[\alpha]{\frac{P_t}{\mathcal{T}_{cs}}} \ . \tag{2.4}$$

2.4.2 Transmit Range

The transmit range of the transmitter X, TR_x, is the area around X, within which a receiver can detect and correctly decode the transmitted signal with high probability, assuming no interference. That is, the signal power level P_r at the receiver is greater than $\mathcal{T}_{rx}$. According to this definition and the large scale propagation models [8], the radius of the TR is

$$R_{tr} \propto \sqrt[\alpha]{\frac{P_t}{\mathcal{T}_{rx}}} \ . \tag{2.5}$$

Further, let γ denote the ratio of $\frac{R_{cs}}{R_{tr}}$. We have

$$\gamma = \sqrt[\alpha]{\frac{\mathcal{T}_{rx}}{\mathcal{T}_{cs}}} \ . \tag{2.6}$$

Because $\mathcal{T}_{rx}$ and $\mathcal{T}_{cs}$ are characteristics of the wireless nodes physical layer implementation, γ can be considered as a constant.

2.4.3 Interference Range

The interference range of the receiver Y, IR_y, is the area around Y, within which another node's simultaneous transmission would corrupt the received signal at Y. That is, within this region another transmission would cause the SINR at Y to fall below $SINR_0$. If all nodes transmit with the same power level, the radius of the IR is

$$R_{ir} = d\sqrt[\alpha]{SINR_0} \ , \tag{2.7}$$

where d is the T-R distance [12]. Depending on d, the IR may be wholly or partly covered by the CSR. Notice that (2.7) does not consider aggregate interference from multiple transmitters.

2.5 Hidden and Exposed Nodes

Hidden and exposed node problems are two essential challenges in wireless MAC protocol design. In Chapter 4, we propose a MAC layer power control scheme based on an investigation of these two issues. Here, we introduce these issues by associating them with the three-range physical model.

2.5.1 The Hidden Node Problem

In general, a hidden node is one that the MAC allows to transmit despite the fact that such a transmission may cause a collision. For CSMA, hidden nodes are the

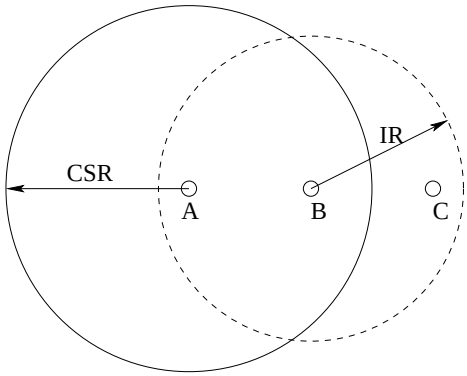


Figure 2.1: The hidden node problem.

ones that are out of the CSR of a transmitter but within the IR of the intended receiver. Consider the case shown in Figure 2.1. A is transmitting to B. Node C, which is far away from A but close to B, can not sense this transmission because it is out of CSR_A. Thus, C determines the channel to be idle, and might commence concurrent transmission corrupting the received signal at B. In this scenario, C is a hidden node. From the perspective of the three-range model, the receiver's IR determines the potential hidden nodes.

The existence of hidden nodes fundamentally reflects a failure on the part of the MAC to actually reserve the spatial resource needed to support a transmission. MAC designs must strive to minimize or better yet eliminate hidden nodes.

2.5.2 The Exposed Node Problem

In general, an exposed node is one which the MAC causes to be silent despite the fact that its transmission would not cause collision. For CSMA, exposed nodes are ones that are within the CSR of a transmitter but out of the IR of the intended receiver. Consider the case shown in Figure 2.2. A is transmitting to B. Node C, which is far away from B but close to A, can sense this transmission because it is within CSR_A. Thus, C determines the channel to be busy and keeps silent, even

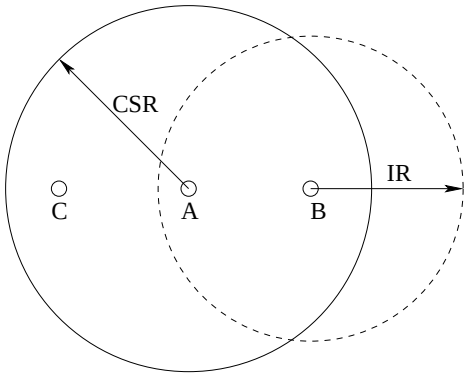


Figure 2.2: The exposed node problem.

though its concurrent transmission would not corrupt the received signal at B. In this scenario, C is an exposed node. From the perspective of the three-range model, the transmitter's CSR determines the potential exposed nodes. The existence of exposed nodes wastes spatial resources, and should be minimized.

2.6 The IEEE 802.11 MAC Protocol

Currently, the IEEE 802.11 MAC protocol [1] is the most widely deployed wireless LAN MAC protocol. It combines CSMA with a collision avoidance scheme based on *Request To Send* (RTS)/*Clear To Send* (CTS) exchange that attempts to solve the hidden and exposed node problems. Even though these problems are not completely eliminated and some other deficiencies exist, they have not prevented the 802.11 MAC from being extremely successful.

The 802.11 MAC works as follows. All transmissions use CSMA and only proceed when the channel is sensed to be idle. For broadcasts, the transmitter simply sends out broadcast messages. Nodes that receive this message do not reply with acknowledgements. For unicasts, reliable transmission is guaranteed by a four-way handshake between the transmitter and the desired receiver. Figure 2.3 illustrates this procedure. Once transmitter A senses the channel to be idle, it sends an RTS to

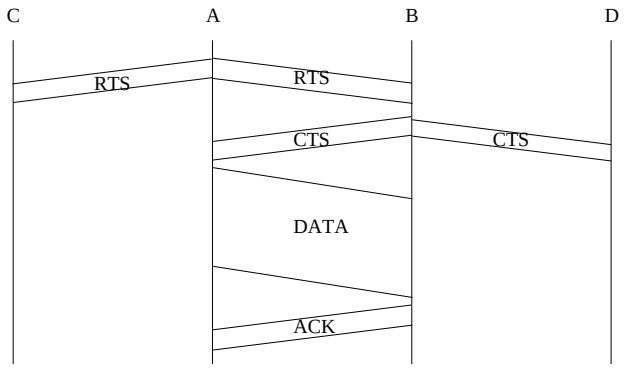


Figure 2.3: A four-way handshake between the transmitter and the receiver.

receiver B which, in turn, replies with a CTS. Meanwhile, the nodes in the vicinity that receive the RTS or CTS defer their own transmissions for a period that is long enough for the subsequent DATA/ACK exchange. When the RTS/CTS exchange completes, A sends data packet and B replies with an ACK if the data packet is correctly decoded. If A does not receive a corresponding CTS or ACK within a predefined time interval, it initiates retransmissions.

Two things determine what nodes remain silent in IEEE 802.11. First, CSMA, in which nodes are silent if the strength of some other transmissions exceeds the carrier sense threshold. Second, nodes are silent if they have decoded either a RTS or CTS, each of which contain information about the duration of another nodes' transmission. This duration information is used to create a "virtual" carrier.

802.11 MAC uses an exponential back-off scheme to avoid and resolve collisions. A node maintains a contention window size, w. After every transmission, it has to wait for m time slots before sensing carrier. Here, m is a number randomly chosen from 0 to w. w is doubled every time when the node sends an RTS but no CTS is received or when the node sends a data packet but no ACK is received. This exponential backoff is repeated until a retry limitation is reached. In this way, collisions could be reduced and every node has a fair chance of accessing the channel.

Two things prevent a node from obtaining the channel. One is that the node may be silenced by other ongoing transmissions. The other is that the node may fail to reserve channel because of collisions. How soon the node is able to gain the channel determines the delay a packet experiences at the node and thus contributes perhaps significantly to queuing latency. When traffic load is high, both channel contention and the chances of collision increase, resulting in increased queuing latency.

2.7 Rate Adaptation

Transmission rate, e.g. data rate, is the number of bits being transmitted during a unit time interval. It is determined by the symbol rate (or channel bandwidth) and the number of bits carried by each symbol (or modulation scheme). Rate adaptation is a technique that optimizes transmission rate according to channel conditions so as to improve spectral efficiency. It can be done by fixing the symbol rate and using different modulation schemes, or by fixing the modulation and changing the symbol rate. In practice, varying the symbol rate is not easy and, therefore, changing the modulation scheme is the common way to adapt transmission rate [14]. In this section, we introduce some basic background knowledge about this technique, which is needed for the discussion in the subsequent chapters.

Shannon's Theorem [15] states that, for a *Additive White Gaussian Noise* (AWGN) channel, the maximum transmission rate (or channel capacity), r_{max}, is a logarithmic function of the signal to noise ratio, SNR, at the receiver,

$$r_{max} = W_c \; log_2(1 + SNR) \tag{2.8}$$

where W_c is the channel bandwidth, P_t is the transmit power, d is the sender-receiver distance, β is a constant, N is the interference-plus-noise, and α is the

path-loss exponent. Rewriting (2.8), we have

$$SNR = 2^{\frac{r_{max}}{W_c}} - 1 \ . \tag{2.9}$$

This shows that the SNR threshold for correct signal reception increases as transmission rate increases.

For a given transmission rate (or modulation scheme), the probability of a symbol error, $\mathcal{P}$, is a monotonically decreasing function of SNR. For example, for a AWGN channel, the error probability for M-array phase modulation is

$$\mathcal{P}_M = 2Q\left(\sqrt{2k\gamma_b}sin\left(\frac{\pi}{M}\right)\right) \ , \tag{2.10}$$

where $Q(x)$ is the Gaussian Q-function, $k = log_2 M$, and γ_b is the SNR of a bit [15]. It can be observed that as M increases, γ_b increases. In another words, the SNR threshold increases as the transmission rate increases.

Previously, we introduced an idealized three-range model which includes the CSR, TR, and IR. Since transmission rate impacts the SINR threshold for signal reception, it impacts the size of the TR and IR. However, transmission rate does not change the carrier sensing threshold, $\mathcal{T}_{cs}$, and, therefore, does not change the size of the CSR.

2.8 SINR Estimation

Our work assumes that the physical layer at the receiver is able to measure the received signal power when it is strong enough to overcome interference and the packet is actually received. This can be achieved using SINR estimation. Suppose that Y is the total power measured at the receiver. We have $Y = P_r + I + N_0$, where P_r is the received signal power, I is the interference, and N_0 is thermal noise. Let $\mathcal{S}$ denote the SINR, e.g. $\mathcal{S} = \frac{P_r}{I+N_0}$. If $\mathcal{S}$ is known, we can obtain P_r as well as $I + N_0$.

More specifically, we have

$$P_r = \frac{\mathcal{S}}{1+\mathcal{S}} Y \tag{2.11}$$

and

$$I + N_0 = \frac{1}{1+\mathcal{S}} Y \ , \tag{2.12}$$

where $\mathcal{S}$ is the measured SINR.

Off-the-shelf wireless cards readily provide such measured values using SINR estimators like the ones discussed in [16]. We briefly introduce several major SINR estimation techniques developed over the last few decades. Notice that, in this section, we assume interference to be AWGN and, when we mention noise, we mean interference-plus-thermal-noise.

Many practical SNR estimators have been developed over the last few decades, including the *Split-Symbol Moments Estimator* (SSME) [17, 18], the *Maximum-Likelihood Estimator for SNR* (ML) [19, 20, 21, 22], the *Squared Signal-to-Noise Variance Estimator* (SNV) [23], the *Second- and Fourth-Order Moments Estimator* (M_2M_4) [24, 25, 26] and the *Signal-to-Variation Ratio Estimator* (SVR) [27]. These estimators derive the SNR from the baseband, sampled, data-bearing received signal. The data may either be known, as is the case for a training symbol, or unknown to the receiver. The techniques that derive the SNR estimates solely from the unknown, information-bearing portion of the received signal are known as in-service SNR estimators and are of particular interest. Pauluzzi *et al.* [16] did extensive experiments to compare the performance of these techniques. They concluded that the best estimator to use depends on the given application. Specifically, if known data is available at the receiver, the ML and SNV estimators perform so well as to make it not worthwhile to find better estimators. If an uninterrupted SNR estimator

is required for an application, then one of the in-service SNR estimators would be more appropriate. The best in-service estimator to use depends on block length, the number of samples per symbol available, the type of modulation used, etc. Finally, the estimators that they have found to to perform best are also relatively easy to implement [16].

Using one of these SNR estimation techniques, the receiver can measure the SNR of a received signal and separate the average signal power from the noise power.

2.9 Signal Power Approximation

If SINR estimation is not used, we can approximate the received signal power to be the total power measured at the receiver, e.g. $P_r \approx Y = P_r + I + N_0$. In this case, it is important to investigate the error margin introduced by such approximations. We investigate this issue under two scenarios restricted to our current work.

The first scenario is one where the receiver experiences low interference during signal reception. We are interested in this scenario because many 802.11-based power control schemes use a control packet exchange, for example RTS/CTS, to estimate channel conditions by measuring the power level of the control packet at the receiver. The IEEE 802.11 standard specifies that the sender sends an RTS only when the channel is idle and that the receiver replies with a CTS only if the channel is idle. Additionally, most physical layers are designed such that, when the receiver senses a signal power that is greater than a threshold ($\mathcal{T}_{cs}$ for IEEE 802.11 standard), it synchronizes with the corresponding transmitter, starts sampling and, therefore, is not able to capture upcoming even stronger signals. Thus, before and after the RTS reception, the receiver must be in an idle state and the interference around is less than $\mathcal{T}_{cs}$. Otherwise, either the RTS is dropped or the CTS is not replied. Moreover, if we assume that the interference is mainly caused by other concurrent transmissions, it is very likely that, during the RTS reception,

the interference-and-noise around the receiver is less than $\mathcal{T}_{cs}$, e.g., $I + N_0 < \mathcal{T}_{cs}$, because no transmissions are shorter than RTS's. Thus, P_r, is bounded by

$$Y - \mathcal{T}_{cs} < P_r < Y \ , \tag{2.13}$$

where Y is the total power measured at the receiver and $Y = P_r + N_0 + I$. If we approximate P_r to be one of the bounds, the error margin, P_δ, is less than $\mathcal{T}_{cs}$. Moreover, to guarantee signal reception, P_r should be greater than $\mathcal{T}_{tx}$. Therefore, the ratio of P_δ to P_r is less than $\frac{\mathcal{T}_{cs}}{\mathcal{T}_{rx}}$, which is about 3% for the parameters given in Table 4.1. In addition, if we approximate P_r to be the average of the two bounds, P_δ is reduced by half.

The second scenario, which is more general, is that the transmitted signal is correctly received at the receiver despite interference. Otherwise, if the signal is corrupted, its power approximation is not needed. To guarantee reliable signal reception, the ratio of P_r and $N_0 + I$ must be greater than the SINR threshold, $SINR_0$, e.g. $\frac{P_r}{N_0+I} > SINR_0$. Because the total power measured at the receiver is $Y = P_r + N_0 + I$, P_r is bounded by

$$Y \frac{SINR_0}{1 + SINR_0} < P_r < Y \ . \tag{2.14}$$

Therefore, P_δ is less than $\frac{1}{1+SINR_0} Y$ and the ratio of P_δ to P_r is less than $\frac{1}{SINR_0}$, which is about 10% for the parameters given in Table 4.1. In addition, if we approximate P_r as the average of the two bounds, P_δ is reduced by half.

Thus, signal power approximation during control packet exchange may be more accurate than that during data packet reception. It is worthwhile to investigate the impact of these error margins on our proposed power control schemes. We leave it as potential future work.

2.10 Positioning Techniques

In Chapter 4 and 6, knowledge of the sender-receiver distance, d, is needed. If we ignore small-scale fading and assume that the path-loss exponent is known, we can derive d from the transmit and receive signal power levels. If we relax these assumptions, d can be obtained using positioning techniques, such as the *Global Positioning System* (GPS) [28] and triangulation [29, 30].

GPS is the only fully-functional satellite navigation system. A constellation of more than two dozen GPS satellites broadcasts precise timing signals by radio, allowing any GPS receiver to accurately determine its location anywhere on Earth. But, GPS works well only in line of sight where satellite visibility is high and it is not efficient for indoor communication systems where the reception of the satellite signal is limited.

The triangulation technique works as follows. First, neighboring nodes estimate distance from each other using transmit and receive signal power levels and establishes relative local coordinates. Thereafter, global coordinates are established for each node within wireless networks. Therefore, the distance between any two nodes can be deduced. The limitation of triangulation is that it has to estimate the sender-receiver distance using transmit and receive power, which is limited by the specific large-scale path loss model, much the same as our design.

2.11 Routing Protocols for Multihop Wireless Networks

At the network level, routing protocols play a key role. Based on when and how routes are discovered and maintained [31], routing protocols in multihop wireless networks can be divided into two categories: table-driven (proactive) and on-demand (reactive) routing. It is generally believed that on-demand routing is more appropriate for multihop wireless networks [31].

2.11.1 Table-Driven Routing Protocols

For table-driven routing protocols, a node attempts to maintain one or more routing tables that contain up-to-date routing information to every other node in the network. The typical examples are Destination-Sequenced Distance-Vector Routing (DSDV) [32], Clusterhead Gateway Switch Routing (CGSR) [33] and Wireless Routing Protocol (WRP) [34]. The areas in which they differ are the number of necessary routing-related tables and the method by which changes in network topology are broadcasted [31].

To discover routes, each node periodically broadcasts to its neighbors its view of the distance to all the other nodes (distance vector) or to the whole network the status of its links (link status). Based on the advertised information, each node computes the path to all the others in the network. To maintain routes, each node periodically broadcasts routing updates to cope with network topology changes.

2.11.2 On-Demand Routing Protocols

For on-demand routing protocols, routes are created only when needed by the source node. Typical examples are Dynamic Source Routing (DSR) [35], Ad Hoc On-Demand Distance Vector Routing (AODV) [36] and Temporally Ordered Routing Algorithm (TORA) [37].

Most of the on-demand routing protocols employ a *blind flooding* technique to discover routes, which works as follows. The source broadcasts a *route request* (RREQ) packet which contains the source address, the destination address, and a unique identification number. The RREQ is flooded throughout the whole network until it reaches the destination or an intermediate node who has the routing information to the destination. During this process, a reverse path from the destination to the source is established. Then, the destination, or the intermediate node, generates a *route reply* (RREP) packet and sends it back to the source. To return the

RREP, the responding node has three choices. First, if it has a route to the source, it can use that routing information. Second, if symmetric links are supported, the responding node can use the reverse route discovered by the RREQ. Third, the responding node can initiate its own route discovery and piggyback the RREP.

The blind flooding technique requires an intermediate node to forward the first arriving RREQ as quickly as possible and discard subsequent duplicates. Thus, the discovered route is the one with the minimum delivery latency of the RREQ. When network traffic is light, forwarding latency dominates the end-to-end packet delivery latency and so the discovered route is the one with the least number of hops, e.g. the shortest-path. However, when network traffic is heavy, queuing latency due to channel contention dominates the end-to-end packet delivery latency and, therefore, the discovered route may not be the shortest-path.

To maintain routes, each forwarding node monitors the status of its next hop and sends a notification to the source once a link failure is detected. A node that receives this link failure notification removes the routes that contain the failure hop. If the route is still desired, the source may initiate a new route discovery.

2.12 Dynamic Source Routing (DSR)

Because our proposed routing protocols are based on DSR, we introduce it in more details. We choose DSR for two reasons. First, on-demand routing protocols are more appropriate for multihop wireless networks than table-driven routing protocols. Second, DSR is a typical on-demand routing protocol and incorporating our routing strategies into other on-demand routing protocols is essentially the same as incorporating the strategies into DSR.

As with most on-demand routing protocols, DSR employs the blind flooding technique to discover routes. In DSR, the RREQ contains a route record field. Every time before a node forwards the RREQ, it adds its address into the record.

Thus, the route record contains the sequence of hops the RREQ traversed. When the destination receives the RREQ, it copies the record into a RREP packet and replies with the RREP to the source.

When the source sends a data packet, it copies the route record into the data packet header and the data packet is forwarded along the hops listed in the record.

2.13 Enhanced Routing Features

Several enhanced routing features are supported by DSR and other on-demand protocols. We introduce those that are needed in our subsequent study.

2.13.1 Multiple Route Replies

Destinations are allowed to response to every arriving RREQ and thus send multiple RREPs to the source. In addition, an intermediate node that knows the route to the intended destination is allowed to reply with a RREP. If the intermediate node knows multiple routes to the same destination, it chooses one according to certain routing strategy.

2.13.2 Route Cache

A source node is allowed to cache multiple routes to the same destination. Before sending a data packet, the source node chooses a route according to certain routing strategy. Most on-demand routing protocols, including DSR, employ the shortest-path routing strategy. That is, the source node would choose the route with the least number of hops.

Once a route is broken, the source node resorts to another cached route without initiating another route discovery. In this way, routing overhead is reduced substantially.

2.13.3 Route Eavesdropping

A node is allowed to obtain route information by eavesdropping on routing packets being transmitted to other nodes. Thus, passively, a node may populate its route cache. These routes may be used by the node itself, or can be used to short circuit route discoveries.

2.14 Route Discovery and Transmit Power

Transmit power plays a key role in the route discovery process. It determines the set of possible neighbors and thus the network connectivity. The transmit power level at which the routing control packets are sent determines the potential next hops, hop distances, and potential routes between sources and destinations. Essentially, the higher the transmit power, the longer the hop-distance, and the stronger the network is connected.

2.15 Shortest-Path Routing

Based on the metric with which the route is chosen, routing protocols in multihop wireless networks can be divided into shortest-path routing, energy-aware routing, QOS-aware routing, etc. We focus on shortest-path routing because most of our subsequent analysis is based on it.

Shortest-path routing attempts to choose a route that contains the least number of hops, and in doing so tends to maximize each hop-distance. Many classic routing protocols are based on this strategy. For example, DSDV, DSR and AODV.

In addition to the basic routing strategy, two features are important in shortest-path routing. The details of the proof are presented in Appendix A.

Lemma 1 *Let R_{tr} denote the transmit range of the power level at which a shortest path is discovered. The average hop-distance of two consecutive hops along the*

shortest path, $\overline{d}_2$, is bounded by

$$0.5R_{tr} < \overline{d}_2 < R_{tr} \ . \tag{2.15}$$

Lemma 1 suggests that, when using shortest-path routing, the average hop distance cannot be shorter than $0.5R_{tr}$. We will show that MAC layer power control schemes favor short hops. Thus, employing shortest-path routing at the network layer may prevent MAC layer power control schemes from gaining much of an advantage.

Lemma 2 *Let L denote the source-destination distance, and R_{tr} denote the transmit range of the power level at which a shortest path is discovered. The number of hops on the shortest path is a function of R_{tr}, denoted by $H(R_{tr})$. We have the following three inequalities*

$$H(R'_{tr}) \leq H(R_{tr}), \qquad (R'_{tr} > R_{tr}) \ , \tag{2.16}$$

$$H(2R_{tr}) \leq \frac{1}{2}H(R_{tr}), \qquad (2R_{tr} \leq L) \tag{2.17}$$

and

$$H(nR_{tr}) \leq \frac{1}{n}H(R_{tr}), \qquad (n = 1, 2, 3, ..., \ nR_{tr} < L) \ . \tag{2.18}$$

Lemma 2 suggests that, for a given source-destination pair, the number of hops increases as the power level decreases. This is a tradeoff that needs to be considered, when we attempt to minimize hop distances and thus minimize the transmit power at each individual hop.

2.16 Related Work

There has not been very much work done directly on spatial usage. But, in a broader sense, spatial usage relates to all the studies on wireless network capacity and throughput. Because we are interested in improving spatial reuse by controlling transmit power, we present related work in two categories: those related to spatial usage, and those related to power control.

2.16.1 Spatial Usage

The studies we are aware of on spatial reuse include [38, 39, 40, 41] and [42]. Characterizing spatial reuse as the ratio between the transmitter-transmitter (T-T) distance and T-R distance, Guo *et al.* [38] showed that increasing transmit power has little effect on spatial reuse in co-channel interference limited scenarios. Zhu *et al.* [39] demonstrated that physical carrier sensing with optimal tuned sensing thresholds improves spatial reuse and network throughput. Motivated by the same considerations, Yang *et al.* [40] took MAC overhead into account and identified the optimal carrier sensing range. Using an enhanced virtual carrier sense, namely *Aggressive Virtual Carrier Sensing* (AVCS), Ye *et al.* [41] improved the performance of IEEE 802.11 based ad hoc networks from the perspective of spatial reuse. Gupta [42] derives an asymptotic upperbound on wireless network capacity by analyzing the spatial usage of transmissions. This is related to part of our work in Chapter 3. But we study the concrete functional form of the bounds on spatial usage, which means we study not only the asymptotic expression but also the constants.

2.16.2 The Border Effect

The border effect is a common issue when analyzing spatial data [43]. In networking systems, it is caused by the fact that the nodes on the border of the network have a different behavior and characteristics from those inside. This causes inconsistences

when comparing a theoretical analysis, which often ignores these effects, with experimental results, which take them into account. The border effect impacts our spatial usage analysis in Chapter 3.

The typical approach to eliminate the border effect is using toroidal geometry. That is, the network topology is modeled such that nodes at the border are considered as being close to nodes on the opposite border. Thus, a flat simulation area becomes a torus [43]. This approach is used in studies such as user distribution and channel assignment in cellular networks [44, 45, 46, 47, 48], multihop wireless networks connectivity analysis [49, 50, 51], and others.

Bettstetter *et al.* [49] proposed another approach to eliminate the border effect. They divided the entire simulation area into disjunct zones: a border zone and an inner zone. Only nodes that are located in the inner zone are considered for the statistics of the simulation. A disadvantage of this approach is that the number of nodes that contribute to the statistics of the simulation decreases as the size of border zone increases. This increases the required simulation time, perhaps significantly.

2.16.3 Power Control

Transmit power has a significant impact on network capacity. Theoretic analysis of multihop wireless network capacity can be found in [42, 52, 53, 54, 55]. Gupta *et al.* [42] derived upper and lower bounds on the capacity of ad hoc networks. Weber *et al.* [52] derived the bounds for transmission capacity of CDMA ad hoc networks and showed that frequency hopped CDMA (FH-CDMA) achieves a higher transmission capacity than direct sequence CDMA (DS-CDMA) for path loss exponent that is greater than 2. Hasan *et al.* [54, 55] showed that the performance of CDMA (both DS-CDMA and FH-CDMA) wireless ad-hoc networks can be improved by employing a suitable sized guard zone around each receiver. With these guard zones,

transmissions are scheduled in a "carrier sensing" manner. They identified the optimal size of the guard zone using stochastic geometry and implemented it using DS-CDMA techniques.

Power control techniques have been used in multihop wireless networks to control network topology and construct energy-efficient broadcast (or multicast) trees [56, 57, 58, 49, 59, 60, 61]. More importantly, power control improves network performance with respect to energy conservation [62, 63, 64, 65, 66, 67, 68, 69, 70, 71], fairness [72, 73], throughput and packet delivery latency [74]. In this dissertation, we focus on its effect on throughput and latency.

Several joint scheduling and power control algorithms have been proposed to improve multihop wireless network throughput [75, 76, 77, 78, 79, 80]. ElBatt *et al.* [75] provided a joint scheduling and power control algorithm working in two phases. In the first phase, a centralized TDMA or TDMA/CDMA scheduling algorithm coordinates independent users' transmissions to eliminate strong levels of interference. In the second phase, a distributed power control algorithm determines the set of powers that could be used by the scheduled users to satisfy their transmissions. Behzad *et al.* [76] improved the scheduling and power control strategy in [75] such that both phases are distributed and operate based on local information exchange. Li *et al.* [77] assumed a TDMA-based wireless ad-hoc network and provided a centralized algorithm of joint power control, scheduling and routing. Wang *et al.* [78] proposed a joint scheduling and power control algorithm in ad hoc networks supporting multicast traffic. They considered an ad hoc network composed of stationary nodes that access the channel by using a TDMA/CDMA scheme. Cruz *et al.* [79] developed an integrated routing, link scheduling and power allocation policy for a multihop wireless network that minimizes the total average power consumption to support minimum average rate requirements per link. Their policy requires time synchronization between transmitters, and requires that channel conditions

remain constant over several time slots. Huang *et al.* [80] proposed a cross-layer optimization framework to jointly design the scheduling, power control and adaptive modulation in TDMA/CDMA wireless ad hoc networks.

The above algorithms adjust transmit power based on a TDMA or CDMA scheduling scheme. However, we are focused on CSMA because it is more appropriate for the "infrastructureless" and "distributed" nature of multihop wireless networks.

Some CSMA-based power control schemes have been proposed [81, 82, 83, 84]. These works introduce either a control channel or busy tones to cooperate transmissions among wireless nodes. Wu *et al.* [81] explored the possibility of combining the concept of power control with RTS/CST-based and busy-tone-based protocols to increase channel utilization. Also, using a separate busy-tone channel, Monks *et al.* [82] presented a power controlled MAC protocol within the collision avoidance framework to increase spatial reuse. In [83], Lin *et al.* enhanced the 802.11 standard by improving the handshaking mechanism and adding a control channel to notify the neighbors around a receiver about the noise tolerance. Thus, the neighbors can adjust their transmission power to avoid packet collision at the receiver. Alizadeh-Shabdiz *et al.* [84] proposed a new IEEE 802.11-based MAC protocol that incorporates multiple transmit power options in the MAC layer by letting nodes adaptively select the transmit power and vary the coverage area. They showed that this MAC protocol reduces network latency when compared to the standard IEEE 802.11 DCF.

The above algorithms employ multiple channels and some may need multiple transceivers at each node, which complicates the hardware design and restrict the applications. To make power control schemes more practical and compatible with the widely used IEEE 802.11 standard, some research groups focus on designing power control algorithms based on single-channel, single-transceiver

systems [85, 86, 87, 88, 7]. Agarwal *et al.* [85] proposed a power control loop for ad-hoc wireless networks, which is similar to those commonly found in cellular CDMA networks. They showed that this power control loop increases overall network throughput by 15%. Poon *et al.* [86] proposed that, after special handling of control packet exchange, data packets are transmitted at the minimum required power level. Although this is a simple strategy, data packets are vulnerable when they are transmitted at the minimum required power level. Yu *et al.* [87] proposed a *Power-Stepped Protocol* (PSP) to enhance spatial utilization in clustered mobile ad hoc networks. The essential idea of PSP is that each node can operate at a different radio power level but no more than one level higher or lower than that of any of its neighbors. The algorithm specifies that a node increases its power when the number of neighbors is less than a predefined threshold and decreases its power when the number of neighbors is greater than a predefined threshold. Thus, PSP is sensitive to node density and is suitable in a clustered topology. The drawback of PSP is that it uses periodic hello messages to exchange neighbor set information and, therefore, may result in substantial overhead. In the framework of 802.11 standard, Zhang *et al.* [88] studied the correlation between the necessary transmit power of RTS, CTS, DATA and ACK packets. The essential idea is to use virtual carrier sense to suppress the transmission of the potential interferers and, therefore, guarantee the next packet reception. For example, the transmission of CTS clears a floor around the receiver so that the upcoming data packets won't be corrupted. Virtual carrier sense works well if all the neighbors in the transmit range decode the transmitted signal correctly and, therefore, keep silent for the rest of the packet exchange. However, it is impossible for all the neighbors to correctly decode the transmitted signal due to kinds of reasons such as collisions, etc. Muqattash *et al.* [7] proposed a rather complicate power control MAC protocol, called POWMAC, which uses an access window (AW) to schedule several concurrent data packet transmissions by using a

series of RTS/CTS exchange. In POWMAC, the transmit power is set such that the received signal power at the receiver is high enough to not only overcome current interference but also allow a certain amount of additions to overcome upcoming concurrent transmissions that might be scheduled in the vicinity. Therefore, POWMAC reserves a larger than needed CSR for an individual transmission. Other transmissions within this larger CSR are scheduled using a series of RTS/CTS exchange in the AW interval. A potential limitation of POWMAC is the AW synchronization among neighboring nodes, especially for those that are two hops away, e.g., the nodes that are inside the CSR but out of TR. These two-hop-away nodes should be able to transmit if the transmit power is lower than that specified in POWMAC.

2.16.4 Multihop Wireless Network Routing Protocols

Routing protocols play a key role in network performance. Different route selection criteria favor different performance metrics. Based on these different route selection metrics, multihop wireless network routing protocols can be divided into several categories: energy-efficient routing, load-sensitive routing, channel-aware routing, link-loss-rate routing, stability-aware routing, shortest-path routing, etc. Energy-efficient routing [89, 90, 91, 92, 93, 94, 95, 96, 97] maximizes battery life by minimizing the energy consumption for delivering each individual data packet. Load-sensitive routing [98, 99, 100, 101, 102] improves network throughput by circumventing hot spots and delivering data packets along under-utilized paths. We introduce these load-sensitive routing protocols in detail in Chapter 7. Both channel-aware routing [103, 104, 105, 106] and link-loss-rate routing [107, 108, 109] attempt to reduce packet loss by choosing routes according to channel (link) conditions or signal strength. Stability-aware routing [110] attempts to reduce the frequency of route outage due to mobility by choosing routes based on the stability of mobile nodes.

Among these routing strategies, shortest-path routing [32, 33, 34, 35, 36, 37] is the fundamental one, because it introduces the basic methods of route discovery and maintenance. In shortest-path routing, all link-costs are identical and, therefore, the discovered routes consist of the least number of hops. When network traffic load is not high, the shortest routes deliver packets with the minimal latency. Most other routing strategies differ from shortest-path routing in the way that they use different metrics to assign the link-cost. We briefly introduce some shortest-path routing protocols to provide a glance at different route discovery schemes used in multihop wireless networks.

Destination-Sequenced Distance Vector (DSDV) [32], *Clusterhead Gateway Switch Routing* (CGSR) [33] and *Wireless Routing Protocol* (WRP) [34] are table-driven routing protocols. They maintain up-to-date routing information from each node to every other node in the network, no matter whether the route is needed or not. DSDV [32] uses a distance-vector routing mechanism. Every node maintains a routing table, where all of the possible destinations and the number of hops to each destination are recorded, and then exchanges routing table updates periodically with its neighbors. According to the exchanged information, each node calculates the routing table based on the classical Bellman-Ford algorithm [111]. CGSR [33] introduces hierarchy into multihop wireless networks by grouping wireless nodes into clusters. Within each cluster, a cluster head is elected using a distributed algorithm. CGSR uses DSDV as the underlying routing scheme. It modifies DSDV by using a hierarchical cluster-head-to-gateway routing approach to route traffic from source to destination. In WRP [34], each node maintains four tables and neighboring nodes inform each other of link changes through the use of update messages. Nodes learn of the existence of their neighbors from the receipt of acknowledgements and hello messages.

Dynamic Source Routing (DSR) [35], *Ad-hoc On-demand Distance Vector*

Routing (AODV) [36] and *Temporally Ordered Routing Algorithm* (TORA) [37] are on-demand routing protocols, which discover a route only when it is needed by the source node. These protocols use the blind-flooding technique to discover routes. In Chapter 7, we show that, although blind-flooding does not strictly discover the shortest-path, it tends to find the path consisting of the least number of hops especially when there are not congested areas. DSR is based on the concept of source routing and it is introduced in the previous section in great detail. In TORA, a source node initiates a route creating process by broadcasting a query (QRY) packet to the desired destination. During the process of route creation, each node updates its "height" metric and links are assigned a direction (upstream or downstream) based on the relative height metric of neighboring nodes. Thus, a directed acyclic graph rooted at the destination is established. AODV establishes a reverse path during the process of forwarding the RREQ and a forward path when the RREP is sent back along the reverse path. Both DSR and TORA can provide multiple routes between a source-destination pair while AODV only provides a single route.

In our work, to increase the efficiency of spatial utilization, we use the mini-hop routing strategy, which chooses routes consisting of a greater number of short hops. Combined with an appropriate MAC layer power control scheme, mini-hop routing increases the efficiency of spatial utilization and may improve network performance significantly. Several combined mini-hop routing and power control schemes are proposed in [112, 113, 114, 115].

ElBatt *et al.* [112] presented a protocol to dynamically determine a connectivity range wherein each node adapts it transmit power so as to only reach a subset of the nodes in the network. The connectivity range contains a predefined number of closest neighbors. This scheme employs a contention-free MAC during the connectivity setup phase, and the connectivity range is chosen not based on network connectivity but on choosing a fixed number of the closest nodes. Muqattash *et*

al. [115] proposed an algorithm to produce spatial-efficient routes while maintaining network connectivity. In [115], each node maintains a connectivity set that contains only the neighboring nodes with which direct communication requires less power than indirect communication via any other node that is already in the set. Thereafter, each node adjusts its transmit power just enough to reach the furthest nodes in the connectivity set and uses this power level to broadcast routing control messages. The advantage of [112] and [115] is that existing routing protocols can be used on top of them without modification. But the disadvantage is that the routes discovered may not be optimal in terms of spatial usage.

Narayanaswamy *et al.* [113] presented a routing strategy where a node maintains multiple routing tables each of which corresponds a power level P_i and the routing table for P_i is constructed by sending and receiving routing control packets at P_i. The optimum power level selected for a node is the smallest one whose routing table has the same number of entries as that of the routing table for the maximum power level. Thus, all nodes use the same optimized transmit power. Kawadia *et al.* [114] improves [113] by considering cluster topologies and claims that a common transmit power is inappropriate in this case. The basic idea of their routing schemes is the same, except that [114] uses a lower transmit power level for intra-cluster communications and a higher transmit power level for inter-cluster communications. It recursively looks up the next hop in lower power level routing tables until gets to the lowest power level routing table at which the next hop is reachable. Unfortunately, this routing strategy has enormous routing traffic overhead.

2.17 Notation

For reference, Table 2.1 lists the major notation used in this dissertation.

Notation	Meaning
P_t	transmit power level
P_r	receive power level
P_{min}	the minimal required transmit power defined by path loss propagation
P_{max}	the maximal transmit power
P_{opt}	power level defined by OTP
P_{eotp}	power level defined by EOTP
d_{ref}	the reference distance
G_t	gain of the transmit antenna
G_r	gain of the receive antenna
h_t	height of the transmit antenna
h_r	height of the receive antenna
λ	wavelength
α	path loss exponent
$SINR_0$	SINR threshold for the capture effect
$\mathcal{T}_{rx}$	power level of receiving threshold
$\mathcal{T}_{cs}$	power level of carrier sensing threshold
TR_x	transmit range of transmitter X
CSR_x	carrier sense range of transmitter X
IR_y	interference range of receiver Y
R_{tr}	radius of the transmit range
R_{cs}	radius of the carrier sense range
R_{ir}	radius of the interference range
γ	ratio of R_{cs} and R_{tr}
Ω	spatial usage metric
n_h	number of hops
n_r	number of retransmissions
a	reserved area of a single transmission
η	spatial reuse factor
S	network area
W	channel capacity
$\mathcal{C}$	network capacity
$\mathcal{P}$	network throughput

Table 2.1: List of notation.

Chapter 3

Spatial Usage and Transmit Power

In wireless networks, multiple wireless nodes share a common communication channel and concurrent transmissions interfere with each other. Signals can be correctly received only when the ratio of its received power level to the interference is above a certain threshold. Therefore, to guarantee reliable signal reception, concurrent transmitters must be kept a certain distance away, and transmitters that are close cannot transmit simultaneously. For this reason, every node has to compete for two kinds of limited resource, space and time, both of which are crucial to potential network throughput. First, efficient spatial usage means more concurrent transmissions can be packed together. Second, efficient time usage means more transmissions can be completed in a unit period. Thus, the concept of "space-time" is needed to analyze potential network throughput.

In this Chapter, we have two objectives.

- First, to find a space-time metric, Ω, that quantifies the idea of space-time usage and that can be connected to network throughput.

- Second, to analyze the impact of transmit power on Ω so as to reveal the impact on potential network throughput.

To achieve the first objective, we claim that Ω must account for all factors contributing to the space-time usage in forwarding a packet, e.g. it is the space-time usage of each individual hop summed over all hops from source to destination. Thus, Ω is a metric consisting of both a network factor, which is the number of hops between source and destination, and a MAC layer factor, which is the space-time usage at each individual hop. There might be multiple retransmissions at each hop, and therefore, the space-time usage at each individual hop is the accumulation of the space-time consumed for each single transmission. Moreover, the intuition behind Ω and potential network throughput can be explained as follows. Consider a network system with area, S, and a communication duration, T. The total space-time resource provided by this system is $S \times T$. This is analogous to the capacity of a container with transversal area, S, and height, T. If each packet needs Ω space-time to be successfully forwarded from source to destination, the potential number of packets that could be forwarded is $\frac{S \times T}{\Omega}$, which is the potential overall network throughput. Because S and T are fixed parameters of the system, Ω determines the potential network throughput, or in another words, Ω is an effective indication of the potential network throughput.

To achieve the second objective, we bound each terms in Ω as a function of transmit power and thus bound Ω as a function of transmit power. Associating the relationship between Ω and potential network throughput, the impact of transmit power on potential network throughput is disclosed.

Further, to validate our theoretical analysis, we performed simulation experiments using NS2 [5, 6]. The initial simulation results turn out not to match our analytical results. The source of this mismatch is the border effect that often occurs when dealing with spatial data. Our study of this effect leads to two approaches to

alleviate or eliminate it. By applying those approaches to the same experiments, we obtain a close match between the experimental results and the theoretical predictions.

All of our studies are performed in both *one dimensional* (1D) space and *two dimensional* (2D) space. Though the analysis of 1D case is straightforward, it models the spatial utilization of an end-to-end transmission. In particular, the behavior of each individual traffic flow in 2D or 3D space is similar to a 1D traffic flow. In addition, studying the 1D case gives us insights into the 2D and 3D cases. Because the concepts, definitions and derivations of spatial usage in 3D space is similar to those in 2D space, we do not present them here.

3.1 System Model

The majority of our analysis assumes that the MAC is based on carrier sensing, i.e., CSMA. We make this assumption for two reasons. First, CSMA is widely used in multihop wireless networks and forms the basis for the IEEE 802.11 DCF MAC. Second, in pure CSMA, the reserved area of a single transmission is exactly the CSR, making our analysis tractable.

We make three assumptions throughout the analysis:

1. Every node transmits at the same power level. This means that the size of CSR and TR are identical for each node's transmission.

2. The path from source to destination is discovered by a shortest-path routing protocol.

3. The channel capacity, modulation, and coding scheme are identical throughout the whole network system. Thus, given the packet length, the transmission duration is identical for each transmission taken from source to destination.

As a reminder, R_{cs} and R_{tr} denote the radius of the CSR and TR respectively. In addition, γ is the ratio of R_{cs} and R_{tr}. As discusses in Section 2.4, γ is determined by the wireless node physical layer implementation. It does not vary with transmit power. Moreover, according to our propagation models, the transmit power P_t and the corresponding R_{tr} are a one-to-one mapping. Thus, bounding Ω and its terms as functions of P_t is equivalent to bound them as functions of R_{tr}. For simplicity, we choose the latter.

3.2 An Upperbound on Network Capacity

We begin by deriving an upperbound for the network capacity of a fixed size wireless network as a function of transmit power. We associate the network capacity with transmit power by way of the spatial usage of a single transmission. In addition to the inherent interest of finding the network capacity, this derivation foreshadows the subsequent analysis of network throughput. Further, it allows us to introduce some key concepts about spatial usage in a straightforward context.

We define the network capacity to be the channel capacity times the number of concurrent transmissions. The network capacity is maximized when all communications are single hop and there are the maximum possible number of concurrent transmissions in the network. Thus, for fixed transmission rate, an upperbound on network capacity is simply the channel capacity, W, times the maximum number of concurrent transmissions, MCT. We define three factors that contribute to MCT: the network area, S, the spatial usage of a single transmission, a, and the upperbound on the spatial reuse factor, η_u.

Definition 1 *Define the network area, S, as the smallest convex polygon that contains all the wireless nodes in the network.*

Definition 2 *Define the spatial usage of a single transmission, a, as the area re-*

served by the transmission within which no other concurrent transmissions occur.

The specific MAC mechanism determines a. For CSMA, a is the size of the CSR, denoted by a_{cs}.

Definition 3 *Define the spatial reuse factor, η, as the average number of concurrent transmitters that reuse a unit area.*

The spatial reuse factor, η, captures the efficiency of spatial utilization. Its upperbound, η_u, determines an upperbound of spatial reuse. To derive η_u, we will introduce a model of concurrent transmissions within which η is maximized, yielding η_u. This model will maximize the network capacity as well.

Given S, a_{cs} and η_u, MCT is

$$MCT = \eta_u \frac{S}{a_{cs}} \ . \tag{3.1}$$

Further, network capacity is maximized when there are no transmission failures and all transmissions are at the full channel capacity W. In this case, the upperbound on network capacity is

$$\mathcal{C}_u = \eta_u \frac{S}{a_{cs}} W \ . \tag{3.2}$$

Notice that (3.2) is an accurate estimation under the assumption that S covers the a_{cs} of all the transmissions. However, this assumption may not be satisfied in real networks due to the fact that S may not be able to cover the a_{cs} of the transmissions at the border. For large networks, this effect is asymptotic neglectable. Therefore, (3.2) is asymptotically legitimate. We study this issue in detail in Section 3.7.

Figure 3.1: Concurrent transmission model in 1D space.

3.2.1 Network Capacity in 1D Space

In 1D space, for CSMA, concurrent transmitters must be separated by at least R_{cs}. Therefore, to pack concurrent transmissions as tightly as possible, transmitters must occur exactly R_{cs} apart, as shown in Figure 3.1. This topology represents the best possible spatial utilization. Note, we assume that there is a receiver within the transmit range of each transmitter. Those receivers do not affect our further analysis, and so they are ignored for now.

For this model, the terms in (3.2) are

S : Is the distance between the leftmost and rightmost nodes.

a_{cs} : Since a transmitter reserves an R_{cs}-length area on its left and right side respectively, $a_{cs} = 2R_{cs}$.

η_u : Here, each R_{cs}-length area is reused by two concurrent transmitters. Therefore, $\eta_u = 2$.

Substituting these terms into (3.2), yields an upperbound on network capacity in 1D space,

$$\mathcal{C}_u^{(1D)} = 2\frac{S}{2R_{cs}}W = \frac{S}{R_{cs}}W = \left(\frac{SW}{\gamma}\right)\frac{1}{R_{tr}} \ . \tag{3.3}$$

This upperbound on network capacity is an inverse function of R_{tr}. It increases as transmit power decreases. The intuition of this equation is simple. When nodes transmit with a lower power level, each single transmission utilizes a smaller area. Thus, more concurrent transmissions can be commenced simultaneously, which leads to a higher network capacity.

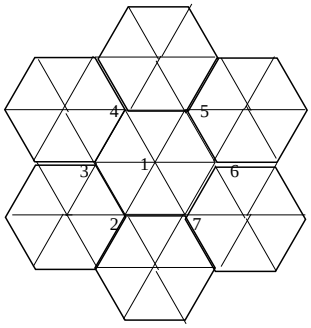


Figure 3.2: Concurrent transmission model in 2D space.

3.2.2 Network Capacity in 2D Space

In 2D space, to maximize network capacity, transmitters must also be placed exactly R_{cs} apart. Following the cellular network structure, some studies [38, 40] proposed a concurrent transmission model that represents the best possible spatial utilization. As shown in Figure 3.2, concurrent transmitters are put at the vertices and the centers of hexagons so that neighboring transmitters are just R_{cs} apart. Again, we assume that there is a receiver within the transmit range of each transmitter. In Figure 3.2, we use hexagons to approximate the circular ranges of the CSR so that we can derive η_u straightforwardly. In addition, we can borrow some research results obtained from such cellular hexagon structures.

For this model, the terms in (3.2) are the following:

S : Is the same as the definition.

a_{cs} : Is the size of the CSR, a circular range around the transmitter with radius R_{cs}. In Figure 3.2, the CSR is approximated by the hexagon. Therefore, $a_{cs} \approx \frac{3\sqrt{3}}{2} R_{cs}^2$.

η_u : The best spatial reuse case is shown in Figure 3.3. Each transmitter reserves a hexagonal area consisting of six equal-lateral triangles. Each equal-lateral triangle is reused by the three concurrent transmitters residing at its vertices.

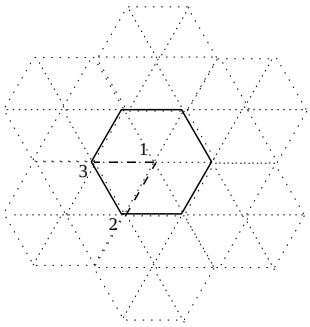


Figure 3.3: A transmitter reserves a hexagon consisting of six equal-lateral triangles each of which is reserved by three concurrent transmitters residing at the vertices.

For example, as shown in Figure 3.3, node 1 reserves a hexagonal area when it is transmitting and, for each of the equal-lateral triangle inside its hexagon, node 1 can share the area with other two concurrent transmitters at the vertices (for example, node 2 and 3). Therefore, each unit area can be reused by three concurrent transmitters for the best case and we have $\eta_u = 3$.

Substituting these terms into (3.2), yields the upperbound on network capacity in 2D space,

$$\mathcal{C}_u^{(2D)} = 3\frac{S}{\frac{3\sqrt{3}}{2}R_{cs}^2}W = \frac{2S}{\sqrt{3}R_{cs}^2}W = \left(\frac{2SW}{\sqrt{3}\gamma^2}\right)\frac{1}{R_{tr}^2}\ . \tag{3.4}$$

This upperbound for the network capacity is an inverse function of R_{tr}^2. It increases as transmit power decreases. When nodes transmit with lower power level, each single transmission utilizes a smaller area. Thus, more concurrent transmissions can be scheduled at the same time, which leads to a higher network capacity. Comparing to (3.3), the network capacity in 2D space varies with the transmit power more aggressively than the network capacity in 1D space.

3.3 A Spatial Usage Metric

Now, we consider the spatial usage of multihop transmissions. We define a metric, Ω, to measure the spatial usage of forwarding a packet from source to destination, and derive its upperbound and lowerbound as a function of the transmit power P_t, or, equivalently, the transmit range, R_{tr}. Because Ω indicates the potential network throughput, our study leads to insights about how power control might be used to improve spatial reuse as well as potential network throughput.

Essentially, Ω is the accumulation of the spatial usage of each individual transmission that a packet undergoes. Thus, not only the number of hops but also the number of retransmissions and the transmission duration at each hop must be taken into account. Ω is defined as below:

Definition 4 *Define a spatial usage metric, Ω, as the total amount of space-time consumed by forwarding a packet from source to destination,*

$$\Omega = \sum_{i=1}^{n_h} n_{ri} a_i t_i \ , \tag{3.5}$$

where n_h is the number of hops from source to destination, and n_{ri}, a_i and t_i are the number of retransmissions, the reserved area, and the transmission duration at the ith-hop respectively.

Note that, if n_h and n_r are set to be one, Ω becomes the space-time usage of a single transmission. In another words, the spatial usage of a single transmission is a special case of Ω.

To find bounds on Ω, we derive bounds on each term individually. A basic assumption is that the transmission duration, t_i, is identical at each hop for a given packet length. In addition, because we are using CSMA and constant transmit power, a_i is identical for each hop, and, as a function of R_{tr}, it is

$$\begin{cases} a_i^{(1D)} = 2R_{cs} = 2\gamma R_{tr} \\ a_i^{(2D)} = \pi R_{cs}^2 = \pi\gamma^2 R_{tr}^2 \end{cases} . \tag{3.6}$$

In the following sections, we will bound the number of retransmissions, n_r, and the number of hops, n_h, as a function of R_{tr}. Combining these will yield the upperbound and lowerbound on Ω.

3.3.1 Hop Distance

We bound the average hop distance, $\overline{d}_h$, along the shortest path as a function of R_{tr}. $\overline{d}_h$ affects two important factors in Ω: the number of hops n_h for a given source-destination (S-D) pair, and the number of retransmissions n_r at each hop. In addition, it is important in localization and distance estimations [116].

Related work is presented by Cheng *et al* [117] and Vural *et al* [116]. Cheng *et al* [117] studied the distribution of the distance to the furthest node in a single broadcast. But this is not the distribution of the single hop distance along the shortest path, because the furthest nodes may not be chosen as the forwarding nodes. Under the assumption of spatially uniform distributed networks, Vural *et al* [116] studied the 1D case and attempted to derive the distribution of the maximal distance in a given number of broadcast cycles. Unfortunately, no closed-form expression is derived. Also, this distribution is not for the multi-hop distance along the shortest path, because the destination may not be in the same direction of the furthest broadcast cycles. Instead of deriving the distribution of the single hop distance, we bound the average hop-distance, $\overline{d}_h$, from the perspective of "shortest-path routing". In particular, we derive bounds that are valid for networks where nodes are arbitrarily distributed.

For reliable transmission, every hop-distance cannot be greater than R_{tr}. Therefore, $\overline{d}_h \leq R_{tr}$. To find the lowerbound on $\overline{d}_h$, we use Lemma 1 (Appendix A)

which shows that the average hop-distance, $\overline{d}_2$, of two consecutive hops is bounded by $0.5R_{tr} < \overline{d}_2 \leq R_{tr}$. Let h denote the number of hops along the shortest path, and $d_{i,i+1}$ denote the hop-distance from node i to $i+1$. We consider two cases: h is even and h is odd.

First, when h is even, we have

$$\begin{aligned} \overline{d}_h &= \tfrac{1}{h}\{(d_{1,2}+d_{2,3})+(d_{3,4}+d_{4,5})+...+(d_{h-1,h}+d_{h,h+1})\} \\ &> \tfrac{1}{h}\{\underbrace{R_{tr}+R_{tr}+...+R_{tr}}_{0.5h}\} = 0.5R_{tr}\ . \end{aligned} \tag{3.7}$$

Second, when h is odd, we have

$$\begin{aligned} \overline{d}_h &= \tfrac{1}{h}\{(d_{1,2}+d_{2,3})+(d_{3,4}+d_{4,5})+...+(d_{h-2,h-1}+d_{h-1,h})+d_{h,h+1}\} \\ &> \tfrac{1}{h}\{\underbrace{R_{tr}+R_{tr}+...+R_{tr}}_{0.5(h-1)}+d_{h,h+1}\} \\ &> \tfrac{1}{h}\{\underbrace{R_{tr}+R_{tr}+...+R_{tr}}_{0.5(h-1)}\} = (\tfrac{h-1}{h})0.5R_{tr}\ . \end{aligned} \tag{3.8}$$

Thus, when h goes to large odd number, $\overline{d}_h$ is also bounded by $0.5R_{tr}$ and R_{tr}.

Combining the above two cases, we know that when h goes to large, $\overline{d}_h$ is bounded by

$$0.5R_{tr} < \overline{d}_h \leq R_{tr}\ . \tag{3.9}$$

Thus, $\overline{d}_h$ is bounded by two linear function of R_{tr}. The larger the R_{tr}, the longer the $\overline{d}_h$. The derivation above applies to both 2D and 1D space. Therefore, (3.9) is valid for both cases.

3.3.2 Number of Retransmissions

Now, we find the upperbound and lowerbound of the number of retransmissions, n_r, at each hop as a function of R_{tr}.

The number of retransmissions, n_r, is a function of the *Packet Error Rate* (PER) p_{per} or *Bit Error Rate* (BER)

$$n_r = \sum_{i=1}^{\infty} i(1 - p_{per})p_{per}^{i-1} = \frac{1}{1 - p_{per}} \ . \tag{3.10}$$

This shows that n_r increases as the PER increases. Moreover, the PER and BER increase as the SINR at the receiver decreases, if we consider the interference to be AWGN. For a given modulation scheme, their relationship can be quantitatively expressed and n_r is a function of SINR, which increases as SINR decreases.

An obvious lowerbound of n_r is 1, since there must be at least one transmission at each hop

$$1 \leq n_r \ . \tag{3.11}$$

To show that n_r has an upperbounded, we derive a lowerbound on the SINR, $SINR_{min}$, under the worst co-channel interference cases, which are modeled by Figure 3.1 and 3.2 for the 1D and 2D space respectively. The number of retransmissions corresponding to $SINR_{min}$ is an upperbound for n_r. Another reason for considering these worst case models is that they approximate network scenarios of high node density and heavy traffic load. It is exactly these situations in which spatial utilization is crucial to overall network throughput.

Our analysis will demonstrate that $SINR_{min}$ is a constant value that does not change as a function of R_{tr}. The intuition behind is that both receive power level and interference vary in the same scale as transmit power varies, resulting in the SINR at the receiver being independent of R_{tr}. Thus, although we will not show

Figure 3.4: The worst SINR in the 1D model

a specific upperbound on n_r and, in fact, the specific upperbound depends on the receiver, we will know it is independent of R_{tr}.

Number of Retransmissions in 1D Space

To facilitate our presentation, we slightly modify the 1D model in Figure 3.1 to be the one in Figure 3.4. All the concurrent transmitters are evenly separated by R_{cs}. A receiver, node 0, is put in the transmit range of node 1. The transmitters on the left of 0 are numbered as odds, and those on the right are numbered as evens. The distance between transmitter 1 and receiver 0 is d ($0 < d \leq R_{tr}$).

Let P_t denote the transmit power corresponding to R_{tr}, α denotes the path loss exponent, β denote the constant for the large-scale propagation, N_0 denotes the thermal noise, and d_i denotes the distance from transmitter i to receiver 0. The SINR at receiver 0 is

$$SINR = \frac{\beta \frac{P_t}{d^\alpha}}{N_0 + \beta \sum_{i=2}^{\infty} \frac{P_t}{d_i^\alpha}} \ . \tag{3.12}$$

From Figure 3.4, it is easy to see that d_i is

$$d_i = \begin{cases} \frac{(i-1)R_{cs}}{2} + d & (i \ is \ odd) \\ \frac{iR_{cs}}{2} - d & (i \ is \ even) \end{cases} \ . \tag{3.13}$$

Substituting d_i into (3.12) and simplifying, we have

$$SINR = \left(N_0 \frac{d^\alpha}{\beta P_t} + \sum_{odd \ i \geq 3} \left(\frac{1}{\frac{(i-1)R_{cs}}{2d} + 1} \right)^\alpha + \sum_{even \ i \geq 2} \left(\frac{1}{\frac{iR_{cs}}{2d} - 1} \right)^\alpha \right)^{-1} \ . \tag{3.14}$$

Obviously, as d increases, each of the three terms increases, and the SINR decreases. Therefore, the SINR is minimized when $d = R_{tr}$, i.e.,

$$\begin{aligned} SINR &\geq \left(N_0 \frac{R_{tr}^{\alpha}}{\beta P_t} + \sum_{odd\ i \geq 3} \left(\frac{1}{\frac{(i-1)R_{cs}}{2R_{tr}} + 1} \right)^{\alpha} + \sum_{even\ i \geq 2} \left(\frac{1}{\frac{iR_{cs}}{2R_{tr}} - 1} \right)^{\alpha} \right)^{-1} \\ &= \left(N_0 \frac{R_{tr}^{\alpha}}{\beta P_t} + \sum_{odd\ i \geq 3} \left(\frac{1}{\frac{(i-1)\gamma}{2} + 1} \right)^{\alpha} + \sum_{even\ i \geq 2} \left(\frac{1}{\frac{i\gamma}{2} - 1} \right)^{\alpha} \right)^{-1} \\ &= \left(N_0 \frac{1}{\beta \mathcal{T}_{rx}} + \sum_{odd\ i \geq 3} \left(\frac{1}{\frac{(i-1)\gamma}{2} + 1} \right)^{\alpha} + \sum_{even\ i \geq 2} \left(\frac{1}{\frac{i\gamma}{2} - 1} \right)^{\alpha} \right)^{-1} \\ &= SINR_{min}^{(1D)} \ . \end{aligned} \tag{3.15}$$

Because both $\mathcal{T}_{rx}$ and γ are constants, and the ratio of R_{tr}^{α} and P_t is a constant according to the large propagation model, $SINR_{min}^{(1D)}$ is a constant that does not change with R_{tr}.

The number of retransmissions $N_r^{(1D)}$ corresponding to $SINR_{min}^{(1D)}$ is the upperbound of n_r. Combining with (3.11) yields

$$1 \leq n_r \leq N_r^{(1D)} \ . \tag{3.16}$$

This shows that n_r is not a function of the transmit range R_{tr}. Instead, it depends on the physical characteristics of the receiver. Note that if R_{cs} is too small such that $\gamma = 1$, $SINR_{min}^{(1D)}$ goes to 0, which makes signal reception impossible and hence a very large number of $N_r^{(1D)}$. In system design, R_{cs} needs to be carefully chosen so that the $N_r^{(1D)}$ is within an acceptable range. A reasonable choice is $R_{cs} > R_{tr} + R_{ir}$, which means that the CSR covers the TR and IR so that no hidden nodes can corrupt the transmission. In this way, a high probability of reliable transmissions can be guaranteed, and the value of $N_r^{(1D)}$ is reasonable.

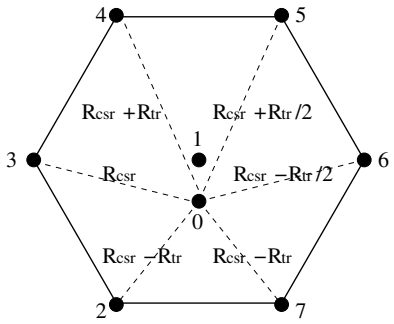


Figure 3.5: The worst SINR in the 2D model

Number of Retransmissions in 2D Space

To facilitate our presentation, we add one receiver, node 0, in the 2D model in Figure 3.2. Node 0 is in the transmit range of transmitter 1, and their distance, d, is constraint by $0 < d \leq R_{tr}$. All the other concurrent transmitters are separated by R_{cs} and located at the vertices of the equal-lateral triangles. Thus, the SINR at receiver 0 is

$$SINR = \frac{\beta \frac{P_t}{d^\alpha}}{N_0 + \beta \sum_{i=2}^{\infty} \frac{P_t}{d_i^\alpha}} = \left(N_0 \frac{d^\alpha}{\beta P_t} + \sum_{i=2}^{\infty} \left(\frac{d}{d_i} \right)^\alpha \right)^{-1} , \tag{3.17}$$

where d_i is the distance from transmitter i to receiver 0.

Note that Figure 3.2 is the cellular telephony structure which has been widely studied. Lee *et al.* [118] showed that the SINR at receiver 0 reaches the lowest level when $d = R_{tr}$ and the distance from receiver 0 to the six first tier transmitters are $R_{cs} - R_{tr}$, R_{cs}, $R_{cs} + R_{tr}$, $R_{cs} + \frac{R_{tr}}{2}$, $R_{cs} - \frac{R_{tr}}{2}$ and $R_{cs} - R_{tr}$ respectively. Figure 3.5 illustrates this scenario. Here, the co-channel interference from other tiers of concurrent transmitters is ignored because it is trivial compared to the interference from the first tier transmitters. Therefore, we have

$$
\begin{aligned}
SINR &\geq \left(N_0 \frac{R_{tr}^{\alpha}}{\beta P_t} + \frac{2}{\left(\frac{R_{cs}}{R_{tr}} - 1\right)^{\alpha}} + \frac{1}{\left(\frac{R_{cs}}{R_{tr}} - 0.5\right)^{\alpha}} + \frac{1}{\left(\frac{R_{cs}}{R_{tr}}\right)^{\alpha}} + \frac{1}{\left(\frac{R_{cs}}{R_{tr}} + 0.5\right)^{\alpha}} + \frac{1}{\left(\frac{R_{cs}}{R_{tr}} + 1\right)^{\alpha}} \right)^{-1} \\
&= \left(N_0 \frac{R_{tr}^{\alpha}}{\beta P_t} + \frac{2}{(\gamma - 1)^{\alpha}} + \frac{1}{(\gamma - 0.5)^{\alpha}} + \frac{1}{(\gamma)^{\alpha}} + \frac{1}{(\gamma + 0.5)^{\alpha}} + \frac{1}{(\gamma + 1)^{\alpha}} \right)^{-1} \\
&= \left(N_0 \frac{1}{\beta \mathcal{T}_{rx}} + \frac{2}{(\gamma - 1)^{\alpha}} + \frac{1}{(\gamma - 0.5)^{\alpha}} + \frac{1}{(\gamma)^{\alpha}} + \frac{1}{(\gamma + 0.5)^{\alpha}} + \frac{1}{(\gamma + 1)^{\alpha}} \right)^{-1} \\
&= SINR_{min}^{(2D)} \ .
\end{aligned}
\tag{3.18}
$$

Because both $\mathcal{T}_{rx}$ and γ are constants, $SINR_{min}^{(2D)}$ is a constant which does not change as a function of R_{tr}.

The number of retransmissions $N_r^{(2D)}$ corresponding to $SINR_{min}^{(2D)}$ is the upperbound of n_r. Combining with (3.11) yields

$$
1 \leq n_r \leq N_r^{(2D)} \ . \tag{3.19}
$$

This shows that n_r is not a function of the transmit range R_{tr}. Instead, it depends on the physical characteristics of the receiver. As in the 1D space, if R_{cs} is too small such that $\gamma = 1$, $SINR_{min}^{(2D)}$ becomes 0 and signal reception is almost impossible. $N_r^{(2D)}$ would be a very large number. In system design, R_{cs} should be carefully chosen so that $N_r^{(2D)}$ is within an acceptable range. Also, a possible choice of R_{cs} is a value greater than $R_{tr} + R_{ir}$, which means that the CSR covers the TR and IR so that no hidden nodes could corrupt the transmission and, therefore, $N_r^{(2D)}$ is reasonably low.

3.3.3 Number of Hops

We find a lowerbound and upperbound on the number of hops, n_h, as a function of R_{tr} for a source, S, and destination, D, separated by a distance L.

A common lowerbound can be found for both 1D and 2D space. If all the forwarding nodes are located on a straight line between S and D, and each hop-distance spans the maximum distance R_{tr}, n_h is minimized, i.e.,

$$\frac{L}{R_{tr}} \leq n_h \ . \tag{3.20}$$

We derive an upperbound for n_h in 1D and 2D space independently.

An Upperbound on the Number of Hops in 1D Space

Since by definition, all nodes lie on a straight line in 1D space, and the minimum average hop-distance is $0.5R_{tr}$, the maximum n_h is $\frac{2L}{R_{tr}}$. Combining with (3.20), we have

$$\frac{L}{R_{tr}} \leq n_h < \frac{2L}{R_{tr}} \ . \tag{3.21}$$

Those bounds show that n_h is an inverse function of R_{tr}, and it decreases as R_{tr} increases. Intuitively, while transmit power increases, each single hop spans a longer distance, and therefore it takes a smaller number of hops from the source to the destination. Notice that when R_{tr} is greater than L, n_h is 1 which means the source can reach the destination in a single hop.

An Upperbound on the Number of Hops in 2D space

Unlike the case in 1D space, forwarding nodes may deviate from a straight line in 2D space, and the upperbound on n_h depends significantly on node distribution. We derive an upperbound for networks where nodes are arbitrarily distributed.

Thereafter, we find a tighter upperbound for the special case of where nodes are uniformly distributed.

For an arbitrary network, there is a critical transmit power P_c below which the source and destination are disconnected. Let R_c denote the critical transmit range corresponding to P_c, and h_c denote the number of hops in the shortest path, $Path_c$, at this power level. Given a power level P_t where $P_t > P_c$, the corresponding number of hops at power level P_t cannot be greater than h_c because at least $Path_c$ is reachable. Denote the number of hops in the shortest path as a function of transmit range R_{tr}, i.e., $H(R_{tr})$. We have $H(R_c) = h_c$.

According to Lemma 2 (Appendix A), we know

$$H(2R_{tr}) \leq \frac{1}{2}H(R_{tr}) \qquad (R_{tr} \geq R_c) \ . \tag{3.22}$$

This is a recurrence inequality which describes a function in terms of its value on smaller inputs. Cormen *et al.* [119] introduces three methods of finding the asymptotic bounds for the recurrence equation or inequality. One of them is the substitution method, which involves guessing the form of the solution and then using mathematical induction to find the constants and show that the solution works. Here, we use the substitution method to find the upperbound of $H(R_{tr})$.

First, we guess the solution is

$$H(R_{tr}) \leq \frac{k}{R_{tr}} \ , \tag{3.23}$$

where k is an appropriately chosen constant.

Second, we show this solution holds for the base case which is $R_{tr} \in [R_c, 2R_c)$. According to Lemma 2, we have

$$H(R_{tr}) \leq H(R_c) = h_c = \frac{2h_cR_c}{2R_c} \leq \frac{2h_cR_c}{R_{tr}} \ . \tag{3.24}$$

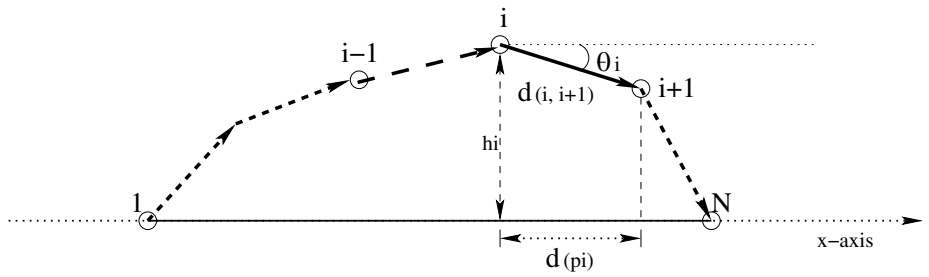


Figure 3.6: The shortest path from source 1 to destination N in a network where nodes are uniformly distributed.

Thus, by choosing $k = 2h_cR_c$, our solution holds for the base case.

Third, we show that if the solution holds for R_{tr}, it holds for $2R_{tr}$. Substituting $H(R_{tr}) \leq \frac{k}{R_{tr}}$ into (3.22) yields

$$H(2R_{tr}) \leq \frac{1}{2}H(R_{tr}) \leq \frac{1}{2}\left(\frac{k}{R_{tr}}\right) = \frac{k}{2R_{tr}} \ . \tag{3.25}$$

Thus, the inductive proof of our solution is completed. Combining with (3.20), we have

$$\frac{L}{R_{tr}} \leq n_h \leq \frac{k}{R_{tr}} \ . \tag{3.26}$$

This equation shows the functional bounds of n_h. That is, given the S-D distance, the number of hops from source to destination is determined by the transmit power level. Intuitively, when transmit power increases, each single hop may span a longer distance, and therefore a smaller number of hops need to be taken from the source to the destination. Note that this bound is valid only when $R_{tr} \leq L$. Otherwise, n_h is always 1.

An Upperbound for Uniformly Distributed Nodes

If the node distribution in the network is known, a tighter upperbound for n_h can be derived. Let us consider a special case where nodes are spatially uniformly

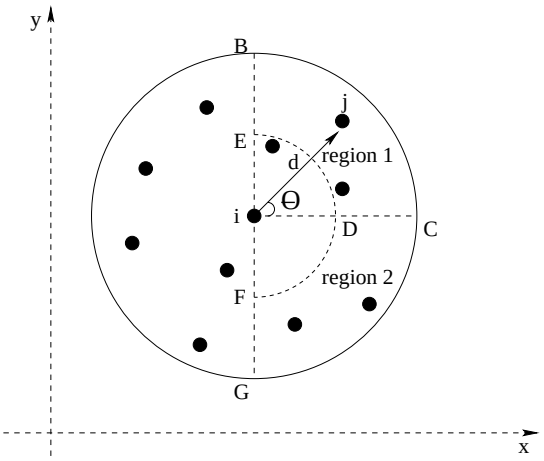


Figure 3.7: Nodes are uniformly distributed in node i's transmit range.

distributed. Figure 3.6 illustrates the shortest path from source 1 to destination N. $i-1$, i and $i+1$ are consecutive forwarding nodes. Without loss of generality, put node 1 at the origin and the S-D direction along the x-axis. The position of each node is determined. Based on these coordinates, we make five definitions:

1. The ith hop is the hop from node i to node $i+1$ ($i \geq 1$).

2. The ith hop distance, $d_{i,i+1}$, is the euclidian distance from node i to node $i+1$.

3. The ith hop direction, θ_i, is the angle between the ith hop and the x-axis.

4. The ith hop progressive distance, d_{pi}, is the projection of $d_{i,i+1}$ on x-axis, i.e., $d_{pi} = d_{i,i+1} cos\theta_i$.

5. The height, h_i, of node i is the perpendicular distance from node i to the x-axis. Specifically, if node i is above the x-axis, $h_i > 0$; if it is on the x-axis, $h_i = 0$; Otherwise, $h_i < 0$.

Consider forwarding node i. Figure 3.7 illustrates its transmit range as a disk with radius R_{tr}. For a node j within the disk, its position can be determined

by the polar coordinates: the distance, d, from i to j, and the angle, θ, between $i \rightarrow j$ and the x-axis. Because nodes are uniformly distributed in the network, they are uniformly distributed in this disk. According to [120], the *probability density function* (PDF) of d and θ are

$$\begin{cases} p(d) = \frac{2d}{R_{tr}^2} & (0 \leq d \leq R_{tr}) \\ \\ p(\theta) = \frac{1}{2\pi} & (-\pi \leq \theta \leq \pi) \end{cases} . \tag{3.27}$$

Now, consider a sub-shortest-path routing (SSPR) protocol which attempts to find a path from the source to the destination, but which is not optimal in terms of the hop count. Note that, in Figure 3.7, we label two regions: BCDE denoted as region 1, and CDFG denoted as region 2. Region 1 is the area within which a node's position satisfies

$$\begin{cases} 0.5R_{tr} < d \leq R_{tr} \\ 0 < \theta < \frac{\pi}{2} \end{cases} , \tag{3.28}$$

and region 2 is the area within which a node's position satisfies

$$\begin{cases} 0.5R_{tr} < d \leq R_{tr} \\ -\frac{\pi}{2} < \theta \leq 0 \end{cases} . \tag{3.29}$$

SSPR proceeds by the following steps. Starting from node 1, each forwarding node i randomly chooses a node from region 1 as its successor if $h_i < 0$, and from region 2 if $h_i > 0$. This is under the assumption that at least one node exists in the desired region. We will analyze the probability of this assumption subsequently. SSPR guarantees that the height of each forwarding node is less than R_{tr} and greater than $-R_{tr}$ and that a path will be found from the source to the destination.

The expected value of a random variable x is defined as

$$E(x) = \int_{-\infty}^{\infty} xp(x)dx \ , \tag{3.30}$$

where $p(x)$ is the probability density function of x. Thus, the expected value of the progressive hop-distance, $E(d_p)$, is

$$E(d_p) = E(d\ cos(\theta)) = E(d)E\left(cos(\theta)\right) \geq 0.5R_{tr}\frac{1}{\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} cos\theta d\theta = \frac{R_{tr}}{\pi} \ . \tag{3.31}$$

Here, d and θ are independent. Thus, the number of hops discovered by SSPR is

$$n_{h-sspr} = \frac{L}{E(d_p)} = \frac{\pi L}{R_{tr}} \ . \tag{3.32}$$

Since shortest-path routing chooses the shortest path from the source to destination, n_h cannot greater than n_{h-sspr}. Combining with the lowerbound of n_h, we have

$$\frac{L}{R_{tr}} \leq n_h \leq \frac{\pi L}{R_{tr}} \ . \tag{3.33}$$

Now, we analyze the probability of successfully finding a SSPR path. Let λ denote the number of nodes in the transmit range disk. For any node, the probability that it will fall into region 1 is

$$p_1 = \int_{0.5R_{tr}}^{R_{tr}} p(x)dx \int_{0}^{\frac{\pi}{2}} p(\theta)d\theta = 0.1875 \ . \tag{3.34}$$

For all λ nodes, the probability that at least one node falls into region 1 is

$$p_\lambda = 1 - (1 - p_1)^\lambda = 1 - 0.8125^\lambda \ . \tag{3.35}$$

The same probability can be obtained for the region 2. Since on average there are $\frac{\pi L}{R_{tr}}$ forwarding nodes along the path, the probability that a SSPR path can be found is that each node has at least one node in the desired region

$$p_{ssrp} = p_{\lambda}^{\frac{\pi L}{R_{tr}}} = \left(1 - 0.8125^{\lambda}\right)^{\frac{\pi L}{R_{tr}}} \quad . \tag{3.36}$$

This shows that, when λ is sufficiently large, $p_{ssrp} \to 1$. In another words, when λ is large, n_h is bounded by (3.33) with a high probability.

3.3.4 The Spatial Usage Metric

Based on the previous studies, we derive the upperbound and lowerbound of Ω. As a summary, we list the bounds for each term:

t_i: This is constant as a function of R_{tr} and is identical at each hop.

a_i: In 1D space, $a_i = 2\gamma R_{tr}$ and, in 2D space, $a_i = \pi\gamma^2 R_{tr}^2$ (See (3.6)).

n_{ri}: It is bounded by $1 \leq n_{ri} \leq N_r^{1D}$ in 1D space and $1 \leq n_{ri} \leq N_r^{2D}$ in 2D space (See (3.16) and (3.19)).

n_h: Given the S-D distance L, n_h is bounded by $\frac{L}{R_{tr}} \leq n_h \leq \frac{2L}{R_{tr}}$ in 1D space and $\frac{L}{R_{tr}} \leq n_h \leq \frac{k}{R_{tr}}$ in 2D space (See (3.21) and (3.26)). In particular, for 2D uniformly distributed networks, $\frac{L}{R_{tr}} \leq n_h \leq \frac{\pi L}{R_{tr}}$ with a high probability when node density is sufficiently high (See (3.33)).

Substituting these terms into (3.5), the spatial usage Ω of forwarding a packet between a L distance S-D pair is bounded by

$$\begin{cases} 2tL\gamma \leq \Omega^{(1D)} \leq 4tL\gamma N_r^{(1D)} & (1D\ space) \\ \\ \left(\pi tL\gamma^2\right) R_{tr} \leq \Omega^{(2D)} \leq \left(\pi tk\gamma^2 N_r^{(2D)}\right) R_{tr} & (2D\ space) \end{cases} . \tag{3.37}$$

This shows that, in 1D space, Ω does not change with R_{tr}, while in 2D space, it increases linearly as a function of R_{tr}. Note that if n_r and n_h are set to be 1, Ω is the spatial usage of a single transmission which has been presented in Section 3.2.

3.3.5 Network Throughput for A Special Traffic Pattern

To confirm (3.37), we consider a special traffic patten and node distribution, bound its potential network throughput using Ω, and compare these theoretical bounds to simulation results.

Consider a distribution in which each traffic flow has the same S-D distance L and uses the same amount of Ω to forward a packet. Let S and T denote the network area and system communication duration respectively, thus the total amount of space-time resource is ST. Because η_u is the maximum spatial reuse factor, the maximum number of packets that the system could forward in T duration is $\eta_u \frac{ST}{\Omega}$. The upperbound on network throughput is

$$\mathcal{P}_u = \eta_u \frac{S}{\Omega} \ (pps) \ . \tag{3.38}$$

This equation is in *packets per second* (pps). Substituting Ω into (3.38) with (3.37), and noting that a packet's transmission duration t is the packet length l divided by channel capacity W. $\mathcal{P}_u$ is bounded by

$$\begin{cases} \eta_u \frac{SW}{4L\gamma N_r^{1D}} \ (bps) \leq \mathcal{P}_u^{(1D)} \leq \eta_u \frac{SW}{2L\gamma} \ (bps) & (\eta_u = 2) \\ \\ \left(\eta_u \frac{SW}{\pi k \gamma^2 N_r^{2D}}\right) \frac{1}{R_{tr}} \ (bps) \leq \mathcal{P}_u^{(2D)} \leq \left(\eta_u \frac{SW}{\pi L \gamma^2}\right) \frac{1}{R_{tr}} \ (bps) & (\eta_u = 3) \end{cases} . \tag{3.39}$$

Thus for this case, the upperbound on network throughput does not change as a function of R_{tr} in 1D space, while it is inversely proportional to R_{tr} in 2D space.

$\mathcal{P}_u$ ignores the overhead, c_o, caused by collisions and protocol control mes-

sages. It can be achieved only when the network area, S, is fully utilized by the traffic flows, which means the spatial reuse factor must achieve η_u. But, in reality, c_o is not neglectable ($0 < c_o < 1$), and the spatial reuse factor η is less than η_u ($0 \leq \eta \leq \eta_u$). Thus, the more realistic bounds are

$$\begin{cases} (1-c_o)\,\eta \frac{SW}{4L\gamma N_r^{(1D)}}\ (bps) \leq \mathcal{P}^{(1D)} \leq (1-c_o)\,\eta \frac{SW}{2L\gamma}\ (bps) \\ \\ \left((1-c_o)\eta \frac{SW}{\pi k\gamma^2 N_r^{(2D)}}\right) \frac{1}{R_{tr}}\ (bps) \leq \mathcal{P}^{(2D)} \leq \left((1-c_o)\eta \frac{SW}{\pi L\gamma^2}\right) \frac{1}{R_{tr}}\ (bps) \end{cases} . \quad (3.40)$$

3.4 Revisiting Our Analysis with Consideration of Rate Adaptation

Power control and rate adaptation are two basic physical layer techniques that might be used to improve the efficiency of spatial utilization. Power control improves spatial utilization by reducing the reserved space for each individual transmission and thus allowing more concurrent transmissions, while rate adaptation improves spatial utilization by reducing the duration of each space reservation and thus allowing more transmissions to be completed in a time period. Thus, the former focuses on "space efficiency" and the latter focuses on "time efficiency".

Our analysis above investigates the relationship between transmit power and network capacity and throughput. It is based on the assumption that every node transmits at the same rate, r, and that r does not change during communications. In this section, we consider a further question. That is, if we allow varying transmission rates, what would the impact be on the above analysis? Further discussion about the impact of rate adaptation on each protocol design is presented in the subsequent chapters.

If we were to consider rate adaptation, two observations are important:

1. First, the size of the CSR is a function of transmit power and the carrier sensing threshold, $\mathcal{T}_{cs}$. Because $\mathcal{T}_{cs}$ does not change with transmission rate, the size of CSR does not change as rate varies.

2. Second, transmission rate impacts the receiving threshold, $\mathcal{T}_{rx}$, and the SINR threshold, $SINR_0$. Therefore, both TR and IR vary with transmission rate.

To introduce rate into our analysis, we need to consider sender-receiver distances. For simplicity, we revise our previous analysis in two steps. In the first step, we assume that all sender-receiver distances are identical and each transmitter adapts its transmission rate to the maximum possible rate, r_{max}, defined by Shannon's theorem. Chapter 2 shows that r_{max} is determined by transmit power P_t and sender-receiver distance d. Because previously we assumed that every sender transmits at the same power, the overall effect is that every sender transmits at the same rate, which is r_{max}. In the second step, we allow sender-receiver distances to vary. Thus, each transmitter may transmit at a different rate.

3.4.1 Network Capacity with Fixed Sender-Receiver Distances

We introduce rate into the analysis of network capacity with the following assumptions:

1. All sender-receiver distances are identical.

2. Every sender adapts its rate to the maximum possible rate, r_{max}, defined by Shannon's Theorem.

Thus, the overall effect is that every node transmits at the same rate r_{max}.

In (3.2), we define the upperbound on network capacity. There are four terms, the network area, S, the size of the CSR, a_{as}, the spatial reuse factor, η_u, and the channel capacity, W. Obviously, changing transmission rate does not change

S, a_{cs} or η_u. The channel capacity, W, is the transmission rate, r_{max}. Therefore, (3.2) can be revised to be

$$\mathcal{C}_u = \eta_u \frac{S}{a_{cs}} r_{max} \ , \tag{3.41}$$

and the upperbounds of network capacity in 1D and 2D space are revised to be

$$\begin{cases} \mathcal{C}_u^{(1D)} = 2\frac{S}{2R_{cs}} r_{max} = \frac{S}{R_{cs}(P_t)} r_{max}(P_t, d) & (1D \ space) \\ \\ \mathcal{C}_u^{(2D)} = 3\frac{S}{\frac{3\sqrt{3}}{2}R_{cs}^2} r_{max} = \frac{2S}{\sqrt{3}R_{cs}^2(P_t)} r_{max}(P_t, d) & (2D \ space) \end{cases} . \tag{3.42}$$

Here, $R_{cs}(P_t)$ and $r_{max}(P_t, d)$ indicate that R_{cs} is a function of P_t and that r_{max} is a function of both P_t and d. Both R_{cs} and r_{max} increase as transmit power increases. Their combined effect is determined by which term changes more significantly as transmit power varies. We briefly analyze the combined effect as follows.

According to Shannon's Theorem, the maximum transmission rate is

$$r_{max} = W_c \ log\left(1 + \frac{P_r}{N}\right) = W_c \ log\left(1 + \frac{\beta P_t}{N d^{\alpha}}\right) \ , \tag{3.43}$$

where W_c is channel bandwidth, N is noise, β is a constant and d is the sender-receiver distance. Substituting (3.43), $R_{cs} = \gamma R_{tr}$ and $R_{tr} = \beta \sqrt[\alpha]{P_t}$ into (3.42), we have

$$\begin{cases} \mathcal{C}_u^{(1D)} = \frac{S}{\gamma R_{tr}} r_{max} = \frac{S W_c log\left(1+\frac{\beta P_t}{N d^{\alpha}}\right)}{\gamma\beta \sqrt[\alpha]{P_t}} & (1D \ space) \\ \\ \mathcal{C}_u^{(2D)} = \frac{2S}{\sqrt{3}\gamma^2 R_{tr}^2} r_{max} = \frac{2S W_c log\left(1+\frac{\beta P_t}{N d^{\alpha}}\right)}{\sqrt{3}\left(\gamma\beta \sqrt[\alpha]{P_t}\right)^2} & (2D \ space) \end{cases} . \tag{3.44}$$

Asymptotically, both $\mathcal{C}_u^{(1D)}$ and $\mathcal{C}_u^{(2D)}$ decrease as P_t increases, which means that network capacity should benefit from reducing transmit power, e.g. reducing

transmit power improves network capacity more significantly than raising transmit rate by increasing transmit power. This conclusion confirms our intuition that, when transmit power varies, the reserved space changes more significantly than transmission duration, because the area reserved by a transmission is linearly proportional to $\sqrt[\alpha]{P_t}$ in 1D space and $\sqrt[\alpha]{P_t^2}$ in 2D space, while the transmission duration (or transmission rate) is a logarithmic function of the transmit power.

However, a key point is that the effect of power control depends significantly on the specific traffic pattern. In another words, to what extent power control improves network capacity depends on how the space conserved by one transmission is utilized by others. Reducing transmit power may not be useful in non-space-limited networks, for example, when the network traffic load is light or the network scale is small. In contrast, rate adaptation improves network performance not only in spatial-limited networks but also in non-space-limited networks.

Therefore, we argue that power control and rate adaptation apply to different network scenarios. It would be an interesting research topic to design a scheme that switches between the two techniques according to the specific network scale and traffic load.

3.4.2 Network Capacity with Varying Sender-Receiver Distances

Now, we allow different sender-receiver distances and, therefore, each sender may transmit at different rate. Suppose that node i transmits at rate r_i. The upperbound on network capacity is

$$\mathcal{C}_u = \sum_{i=1}^{M} r_i = M\left(\frac{1}{M}\sum_{i=1}^{M} r_i\right) = M\overline{r}\ , \tag{3.45}$$

where M is the maximum number of concurrent transmissions and $\overline{r}$ denotes the average transmission rate of all the concurrent transmissions. Substituting $M = \eta_u \frac{S}{a_{cs}}$ into the above equation, we have

$$\mathcal{C}_u = \eta_u \frac{S}{a_{cs}(P_t)} \overline{r}(P_t, d) \ . \tag{3.46}$$

Here, $a_{cs}(P_t)$ and $\overline{r}(P_t, d)$ indicate that both a_{cs} and $\overline{r}$ are functions of P_t. The functional form above is similar to (3.41). The combined effect of a_{cs} and $\overline{r}$ would be similar to that in (3.41).

3.4.3 Spatial Usage of Multihop Transmissions with the Same Hop-Distances

We introduce rate into the spatial usage analysis of delivering a packet from source to destination with the following assumptions:

1. During route discovery process, there is no rate adaptation. Every node uses a basic rate, r_0, to transmit routing control packets. This matches current practice.

2. Each hop-distance is identical.

3. Every sender along the path adapts its data packet transmission rate to the maximum possible rate, r_{max}, defined by Shannon's theorem.

Thus, the overall effect is that the data transmission rate at each hop is the same.

In (3.5), we define the spatial usage for delivering a packet from source to destination. There are four terms, the CSR of each transmission, a, the transmission duration, t, the number of hops from the source to the destination, n_h, and the number of retransmission at each hop, n_r. Transmission rate does not impact a but it impacts the other three factors: n_h, t and n_r.

First, we consider the relationship between the hop distance, d_h, and transmission rate. The hop-distance is determined during route discovery process and, at this stage, every node uses r_0 to transmit routing control packets. Because the

receiving threshold, $\mathcal{T}_{rx}$, varies with r_0, the radius of the TR is a function of both transmit power P_t and r_0, denoted as $\mathcal{R}_{tx}(P_t, r_0)$. The bounds on d_h in (3.9) become

$$0.5\mathcal{R}_{tr}(P_t, r_0) < d_h \leq \mathcal{R}_{tr}(P_t, r_0) \ . \tag{3.47}$$

Thus, the bounds for the number of hops, n_h, in (3.21) and (3.26) are revised to be

$$\begin{cases} \frac{L}{\mathcal{R}_{tr}(P_t,r_0)} \leq n_h \leq \frac{2L}{\mathcal{R}_{tr}(P_t,r_0)} & (1D \ space) \\ \\ \frac{L}{\mathcal{R}_{tr}(P_t,r_0)} \leq n_h \leq \frac{k}{\mathcal{R}_{tr}(P_t,r_0)} & (2D \ space) \end{cases} . \tag{3.48}$$

Second, we consider the relationship between t and the transmission rate. Because data packets are transmitted at r_{max}, we have $t = \frac{l}{r_{max}}$, where l is the packet length.

Finally, we consider the upper and lower bounds on n_r as a function of r_{max}. Because the CSR does not change with r_{max}, the worst case SINR scenarios in Figure 3.4 and 3.5 are still valid, except that the maximum distance between sender 1 and receiver 0 is $\mathcal{R}_{tx}(P_t, r_0)$. The minimum SINR, $SINR_{min}$, derived in (3.15) and (3.18) are valid, except that both $\mathcal{T}_{rx}$ and γ are functions of r_{max}, since γ is defined as the ratio of R_{cs} and R_{tx}, which is equivalent to the ratio of $\mathcal{T}_{rx}$ and $\mathcal{T}_{cs}$. Therefore, $SINR_{min}$ is a function of r_{max}. Another key point is that the maximum number of retransmissions, N_r, is a function of $SINR_{min}$ and the function is determined by the different modulations schemes, e.g. transmission rate r_{max}. As a conclusion, N_r is a function of r_{max}, denoted as $\mathcal{N}(r_{max})$. Thus, (3.16) and (3.19) are revised to be

$$\begin{cases} 1 \leq n_r \leq \mathcal{N}^{(1D)}(r_{max}) & (1D \ space) \\ \\ 1 \leq n_r \leq \mathcal{N}^{(2D)}(r_{max}) & (2D \ space) \end{cases} . \tag{3.49}$$

Substituting the above terms into (3.5), we obtain the spatial usage bounds for delivering a packet from source to destination

$$\begin{cases} 2lL\frac{\Gamma(r_{max})}{r_{max}} \leq \Omega^{(1D)} \leq 4lL\frac{\mathcal{N}(r_{max})\Gamma(r_{max})}{r_{max}} \\ \\ \left(\pi lL\frac{\Gamma^2(r_{max})}{r_{max}}\right)\mathcal{R}_{tr}(P_t, r_0) \leq \Omega^{(2D)} \leq \left(\pi lk\frac{\mathcal{N}(r_{max})\Gamma^2(r_{max})}{r_{max}}\right)\mathcal{R}_{tr}(P_t, r_0) \end{cases} , \quad (3.50)$$

where l is the packet length in bits. Notice that we replace R_{cs} with $\Gamma(r_{max})\mathcal{R}_{tr}(P_t, r_0)$, where $\Gamma(r_{max})$ represents γ, which is a function of r_{max}. Because both R_{tr} and r_{max} are functions of transmit power, the equation shows that the spatial usage is determined by the combined effect of r_{max} as well as the specific functional forms of $\mathcal{R}_{tr}(P_t, r_0)$, $\mathcal{N}(r_{max})$ and $\Gamma(r_{max})$. The exact functional forms depend on the specific physical layer employed, and so we leave the analysis of this combined effect for future work.

3.4.4 Spatial Usage of Multihop Transmissions with Different Hop-Distances

Now, we relax the second assumption above and allow different hop-distances and thus different data packet transmission rates at each hop.

First, we consider the relationship between average hop distance, $\overline{d_h}$, and r_0. For the same reason stated above, we have

$$0.5\mathcal{R}_{tr}(P_t, r_0) < \overline{d}_h \leq \mathcal{R}_{tr}(P_t, r_0) \ . \quad (3.51)$$

Therefore, the number of hops, n_h, is still bounded by (3.48).

Second, we consider the relationship between transmission duration and transmission rate. Let r_i denote the data packet transmission rate at hop i. The transmission duration at hop i is $t_i = \frac{l}{r_i}$, where l is packet length.

Finally, using the same method above, the number of retransmissions at hop i is

$$\begin{cases} 1 \leq n_{ri} \leq \mathcal{N}^{(1D)}(r_i) & (1D\ space) \\ \\ 1 \leq n_{ri} \leq \mathcal{N}^{(2D)}(r_i) & (2D\ space) \end{cases} . \tag{3.52}$$

The spatial usage of delivering a packet from source to destination is defined by (3.5), e.g.

$$\Omega = \sum_{i=1}^{n_h} (n_{ri} a_i t_i) = n_h a \left(\frac{1}{n_h} \sum_{i=1}^{n_h} (n_{ri} t_i) \right) = n_h a \overline{(n_r t)} , \tag{3.53}$$

where $\overline{(n_r t)}$ is the average of $(n_r t)$ over all the hops. Substituting the terms into this equation, we obtain the bounds for spatial usage

$$\begin{cases} 2lL\overline{\left(\frac{\Gamma(r)}{r}\right)} \leq \Omega^{(1D)} \leq 4lL\overline{\left(\frac{\mathcal{N}(r)\Gamma(r)}{r}\right)} \\ \\ \pi lL\overline{\left(\frac{\Gamma^2(r)}{r}\right)}\mathcal{R}_{tr}(P_t, r_0) \leq \Omega^{(2D)} \leq \pi lk\overline{\left(\frac{\mathcal{N}(r)\Gamma^2(r)}{r}\right)}\mathcal{R}_{tr}(P_t, r_0) \end{cases} . \tag{3.54}$$

This functional form is similar to (3.50). It shows that the spatial usage is determined by the combined effect of $\overline{r}$, $\mathcal{R}_{tr}(P_t, r_0)$ as well as the specific functional forms of $\mathcal{N}(r)$ and $\Gamma(r)$. For the same reasons as before, we leave the analysis of this combined effect for future work.

3.5 Observations

Before proceeding to the subsequent experimental validation, we conclude above analysis with several observations.

First, (3.3) and (3.4) show that network capacity can be increased by reduc-

ing transmit power of each individual transmission. Specifically, in 1D space, the network capacity increases linearly and, In 2D space, network capacity increases quadratically as transmit range decreases. Thus, MAC layer power control is an effective way of improving network capacity, which motivates our investigation presented in Chapter 4 and 5.

Second, the power level of each individual transmission cannot be deliberately reduced because it is restricted by sender-receiver distances. Although we may reduce the sender-receiver distances by delivering data packets along routes consisting of short distance hops, this may result in a greater number of hops and perhaps more spatial usage for delivering each packet from source to destination. Fortunately, (3.37) shows that, in 2D space, the spatial usage of delivering each individual packet can be reduced if the packet is forwarded along short hops and if MAC layer power control is employed at each hop along the route. In another words, MAC layer power control combined with a routing strategy that discovers routes consisting of short distance hops may increase potential network throughput. This motivates our investigation on mini-hop routing in Chapter 6.

Third, to achieve the maximum potential network throughput, the spatial reuse factor, η, mush achieve its upperbound, e.g., network space must be fully utilized. To do this, we should circumvent any congested areas and route traffic flows into under-utilized space. This motivates our investigation on load-sensitive routing in Chapter 7.

3.6 Experimental Validation

We use simulations to validate (3.40) in both 1D and 2D space. In 1D space, we measure the throughput of a chain traffic flow as a function of R_{tr}. In 2D space, we measure the throughput of multiple parallel traffic flows as a function of R_{tr}.

Our simulations are performed using ns2 [5, 6]. The MAC layer protocol

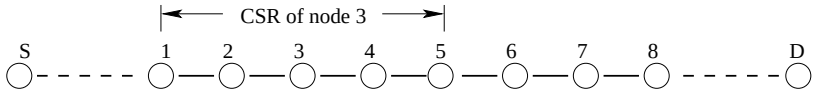


Figure 3.8: A chain traffic flow in 1D space

is the IEEE 802.11 DCF MAC. Because our previous investigation are based on a pure CSMA mechanism, we set $\gamma = 2$ so that the area reserved by "virtual" carrier sensing (RTS and CTS) is always part of the CSR. Moreover, we set the channel capacity to be $1Mbps$, and estimate 802.11 protocol overhead by measuring the throughput of a single S-D pair whose distance is less than the transmit range. This gives us $c_o = 0.3$, which is the minimal overhead caused by the MAC protocol alone, and does not contain the overhead due to routing protocols and collisions etc. The shortest-path routing protocol is DSDV [32]. We increase the periodic update interval in DSDV to 60 seconds so that the routing protocol overhead is negligable.

3.6.1 A Chain Traffic Flow

We use a chain traffic flow to validate (3.40) for 1D space networks. Figure 3.8 shows the network topology where node S and D are the source and the destination respectively. Their distance is $250m$. S sends CBR traffic to D at full channel capacity. We measure the network throughput under different transmit power levels P_k ($k = 1, 2, ..., 8$ and $P_1 > P_2 > ... > P_8$). P_k is set such that its transmit range $R_k = 250m/k$. 50 nodes are uniformly randomly distributed between S and D so that they are connected even when the nodes transmit with the lowest power level P_8. According to this simulation setup, the terms in (3.40) are:

S: It is the distance between the leftmost and rightmost node, which is $S = 250m$.

W: Channel capacity W is set to be $1Mbps$.

L: The S-D distance is $250m$.

γ: γ is set to be 2.

c_o: c_o is measured to be 0.3.

η: In 1D space, the maximum spatial reuse factor is 2. But this is difficult to achieve in real networks. We demonstrate in Appendix C that the spatial reuse factor for a chain traffic flow is approximately 1.33.

N_r^{1D}: Because ns2 only considers the hidden node's interference and signal corruption is based on a "1/0" decision, we estimate the upperbound of retransmissions as follows. Consider Figure 3.8. Assume that node 3 is transmitting, node 4 is receiving, and they are separated by R_{tr} for the worst case. In our simulation, the path loss exponent α and the SINR threshold are set to be 4 and 10 respectively. Thus, according to (2.7), node 4's hidden nodes are those that are less than $1.78R_{tr}$ away. Because node 5 is in the CSR of node 3, it must keep silent. The possible hidden node is node 6 which is two hops away from node 4 and their distance is between R_{tr} and $2R_{tr}$. Here, we ignore node 7 because it is three hops away from node 4 and their distance is from $1.5R_{tr}$ to $3R_{tr}$. While node 4 is receiving, the chances for node 6 transmitting simultaneously is at most $\frac{1}{3}$, because it, at least, has to compete with node 7 and 8. If the hop-distance is smaller, node 6 may even have to compete with node 9. This means that the receiving signal at node 4 has a $\frac{1}{3}$ chance of being corrupted. Substituting p_{per} with $\frac{1}{3}$ into (3.10), we know, in the worst case, the number of retransmissions is approximately 1.5.

Substituting the these terms into (3.40), the potential network throughput for this specific chain traffic flow is bounded by

$$0.76 \times 10^5 \ (bps) \leq \mathcal{P}_{chain} \leq 2.28 \times 10^5 \ (bps) \ . \tag{3.55}$$

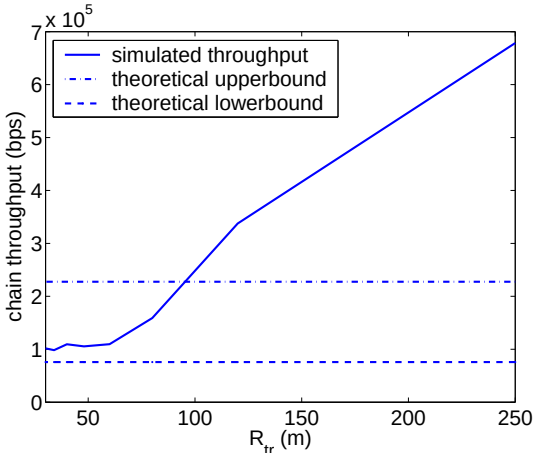


Figure 3.9: Throughput for the chain traffic flow.

Simulation results are shown in Figure 3.9. The x-axis is the transmit range R_{tr} and the y-axis is the throughput in *bps*. The solid line is the experimental result. Two dotted lines are the upper and lower bounds plotted according to (3.55). Unlike our prediction, the simulation curve increases as R_{tr} increases, and it greatly exceeds the estimated upperbound. By studying the previous analysis, we find that this mismatch is caused by the border effect which has been ignored during the derivation of (3.40). We will discuss this issue in detail in the next section.

3.6.2 Multiple Parallel Traffic Flows

We use multiple parallel traffic flows to validate (3.40) for 2D space networks. Figure 3.10 shows the network topology where multiple S-D pairs are evenly spaced in an area of $500m \times 1000m$. The distance between two neighboring traffic flows is d_{tf} which is slightly greater than the carrier sense range R_{cs}. Thus, d_{tf} changes with transmit power, and those traffic flows can transmit concurrently. Moreover, all S-D pairs are of length $500m$, and all sources send CBR traffic at full channel capacity. We measure the network throughput under different transmit power levels P_k ($k = 1, 2, .., 6$, and $P_1 > P_2 > ... > P_6$). The corresponding transmit ranges R_k

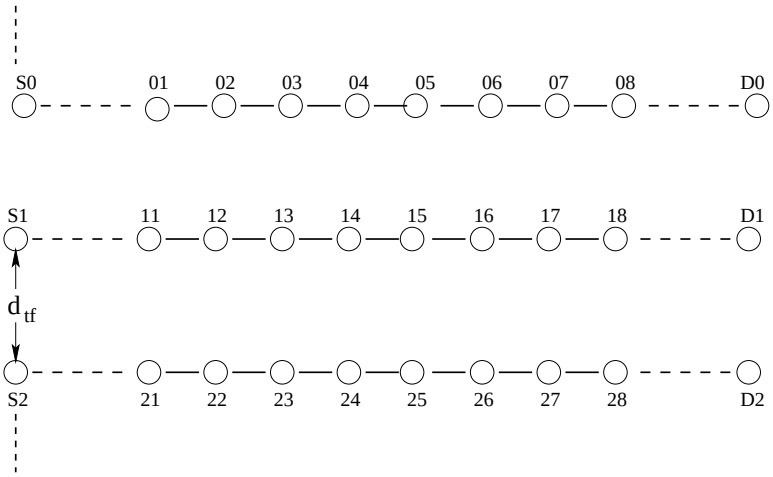


Figure 3.10: Multiple parallel traffic flows in 2D space

$(k = 1, 2, ..., 6)$ are 250m, 160m, 120m, 100m, 80m and 60m respectively. We put enough uniformly distributed nodes along the straight line between each S-D pair so that they are connected even when the nodes transmit at P_6. According to this simulation setup, the terms in (3.40) are

S: According to the setup, $S = 5 \times 10^5 m^2$.

W: Channel capacity W is set to be $1Mbps$.

L: All the S-D distances are $500m$.

γ: γ is set to be 2.

c_o: c_o is measured to be 0.3.

k: For an arbitrary network in 2D space, k is $2h_cR_c$. But, in our simulation setup, nodes are distributed along the straight line from the sources to destinations so that each traffic flow can be considered as forwarding packets in 1D space. Therefore, for this distribution, $k = 2L$.

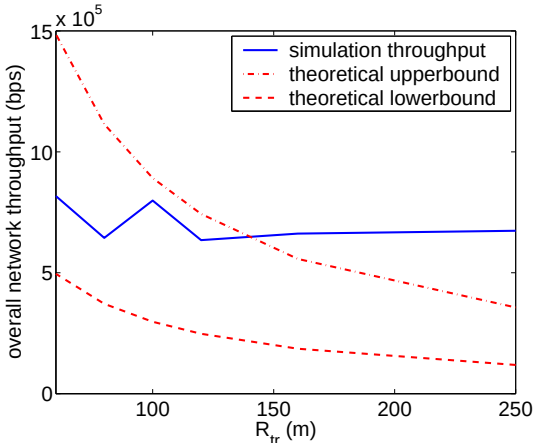


Figure 3.11: Throughput for the multiple parallel traffic flows.

η: The maximum spatial reuse factor in 2D space is 3. But it is hard to achieve in a real network. We demonstrate in Appendix D that η is approximately 1.6 for this simulation setup.

$N_r^{(2D)}$: For the same reason as the chain traffic flow, the upperbound on retransmissions is approximately 1.5.

Substituting those terms into (3.40), the potential network throughput of those parallel traffic flows is bounded by

$$\frac{3.0 \times 10^7}{R_{tr}} \ (bps) \leq \mathcal{P}_{multi} \leq \frac{8.9 \times 10^7}{R_{tr}} \ (bps) \ . \tag{3.56}$$

Simulation results are shown in Figure 3.11. The x-axis is the transmit range R_{tr} and the y-axis is the throughput in *bps*. The solid line is the experimental result. Two dotted lines are the upper and lower bounds plotted according to (3.56). Again, the simulation curve does not match the theoretical estimations, and this mismatch is caused by the border effect which has been ignored during the derivation of (3.40).

3.7 The Border Effect

The border effect occurs often when dealing with spatial data [43]. It is caused by the fact that, in a real network, the nodes on the border of the network have a different behavior and characteristics than those inside. In order to simplify our theoretical analysis, we ignored this fact and obtained asymptotic results that are difficult to validate through experiment because the experiments are dominated by the border effect. For example, in our analysis, the network capacity ((3.3), (3.4)) and the network throughput ((3.39), (3.40)) are derived under the assumption that the network area, S, covers all nodes' CSRs. However, this assumption cannot be satisfied in a real network. Since S is defined to be the smallest convex polygon that contains all nodes, it is not able to cover the CSR of the border nodes. Therefore, all those theoretical equations are asymptotic and could not match our experiments which are dominated by the border effect. To obtain more insight, we re-examine the concurrent transmission models and the network capacity equations proposed in Section 3.2.

Figure 3.12 illustrates the spatial usage of the 1D model in Figure 3.1. Four nodes are evenly separated by R_{cs} and located at position B, C, D and E. S is the area surrounded by the leftmost node 3 and rightmost node 4, which is $3R_{cs}$. According to (3.3), the network capacity is $\mathcal{C}^{(1D)} = 3W$. But we observe that the network capacity is $4W$ since all four nodes can transmit concurrently. This mismatch is caused by the fact that, although node 1 and node 2's CSRs are covered by S, part of node 3 and 4's CSRs are outside of S, as shown in Figure 3.12.

Figure 3.13 illustrates the spatial usage of the 2D model in Figure 3.2. If we only consider the first tier nodes, S is the area of the first tier hexagon, which is $\frac{3\sqrt{3}}{2}R_{cs}^2$. According to (3.4), the network capacity is $\mathcal{C}^{(2D)} = 3W$. But we observe that the network capacity is $7W$ since all seven nodes can transmit concurrently. This mismatch is caused by the fact that, although node 1's CSR is covered by S,

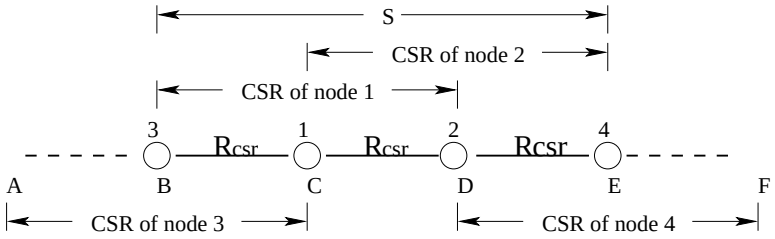


Figure 3.12: The border effect in the 1D model

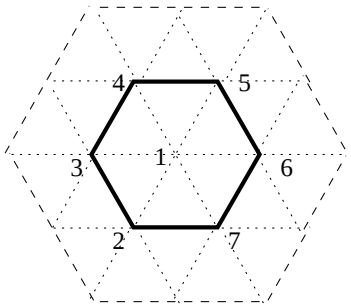


Figure 3.13: The border effect in the 2D model

the CSRs of the six nodes on the border exceed the range of S and go to the second tier hexagon.

Thus, the border effect causes mismatches between theoretical analysis and experimental results. To effectively validate our theoretical analysis, we must eliminate it from our experiment. In the following sections, we use two well known approaches to achieve this. One is to enlarge the network scale and the other is using toroidal geometry [43].

3.7.1 Enlarge Network Scale

When the network scale increases to a sufficiently large value, the error caused by ignoring the border effect is neglectable. Thus, the border effect is alleviated. To have a better understanding, we define network scale and, thereafter, use two examples to explain this argument.

Definition 5 *Define network scale as* $\frac{S}{a_{cs}}$ *where* S *is network area as defined above and* a_{cs} *is the size of the CSR.*

First, enlarge the 1D network in Figure 3.12 so that the network scale $\frac{S}{a_{cs}} \gg 1$. Thus, the potential number of concurrent transmitters $m \gg 1$. The network area, S, surrounded by the outmost nodes is $(m-1)R_{cs}$, and the derived network capacity is $\mathcal{C}^{(1D)} = (m-1)W$. Meanwhile, we observe that the network capacity is mW because m nodes can transmit concurrently. Since $m \gg 1$, $(m-1)W$ is an accurate approximation to mW.

Second, enlarge the 2D network in Figure 3.13 so that the network scale $\frac{S}{a_{cs}} \gg 1$. Therefore, there are m potential tiers of concurrent transmitters where $m \gg 1$. The network area, S, surrounded by the outmost tier is $\frac{3\sqrt{3}}{2}m^2R_{cs}^2$, and the derived network capacity is $\mathcal{C}^{(2D)} = 3m^2W$. Meanwhile, we observe that the network capacity is $3m(m+1)W + W$ because there are $3m(m+1)+1$ nodes that can transmit concurrently. Since $m \gg 1$, the derived network capacity, $3m^2W$, is an accurate approximation to the observed one, $3m(m+1)W + W$.

3.7.2 Toroidal Geometry

In some cases, it is not feasible to simulate a very large scale network which has a wide area and high node density. For example, ns2 [5, 6] is very slow when simulating a network with several hundred nodes. Therefore, we use wrapped around graph to eliminate the border effect, e.g. applying the toroidal geometry.

Again, consider the 1D network in Figure 3.12. The mismatch between the theoretical capacity $3W$ and the experimental capacity $4W$ is caused by the fact that part of node 3 and 4's CSRs are out of S. Imagine that we wrap the chain into a ring so that segment AB overlaps segment DE and segment EF overlaps segment BC, as illustrated by Figure 3.14. All the CSRs are constrained in S. Node 3 and 4 are close to each other and they can not transmit simultaneously. Thus, the

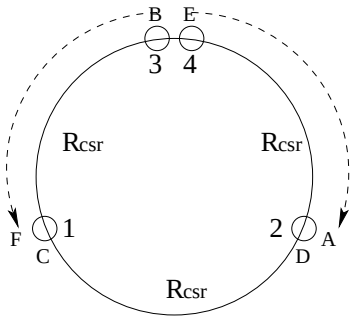


Figure 3.14: Carrier sensing based on toroidal distance

observed network capacity becomes $3W$ which is the same as the theoretical value. In this case, the distance between node 3 and node 4 is not the euclidian distance. Instead, it is the toroidal distance [43, 49] where nodes on the left of the network are close to those on the right.

The same method works for the 2D network in Figure 3.13. The mismatch between the theoretical capacity $3W$ and the observed capacity $7W$ is caused by the fact that part of the CSRs of the boundary nodes are out of S. Imagine that we warp the first tier hexagon into to a globe so that nodes on the top of the network are close to those on the bottom, and nodes on the left are close to those on the right. The distance between two nodes is not the euclidian distance but the toroidal distance. Thus, node 2 overlaps with node 4 and they cannot transmit concurrently. Neither can node 3, 6 and node 5, 7. All the CSRs are constrained in S. The observed network capacity is about $4W$ which is much closer to the derived capacity $3W$ than the previously observed $7W$. It is worth mentioning that the mismatch between the observed $4W$ and the derived $3W$ is caused by the fact that we implicitly assume that S is a rectangle with width S_W and height S_H, which is easy to simulate. But theoretically we use the hexagon to model the best case concurrent transmission. That is, we simulate an area which is larger than the hexagon, $\frac{3\sqrt{3}}{2}R_{cs}^2$, and the observed network capacity is larger than the derived one.

Motivated by those insights, we modified ns2 so that carrier sensing is based on toroidal distance instead of euclidian distance. Let $d((x_1, y_1), (x_2, y_2))$ denote the euclidian distance between two points (x_1, y_1) and (x_2, y_2). The toroidal distance between the two points is

$$\begin{aligned} d_T((x_1, y_1), (x_2, y_2)) = min\{&d((x_1, y_1), (x_2, y_2)), \\ &d((x_1 + S_W, y_1), (x_2, y_2)), \\ &d((x_1 - S_W, y_1), (x_2, y_2)), \\ &d((x_1, y_1 + S_H), (x_2, y_2)), \\ &d((x_1, y_1 - S_H), (x_2, y_2)), \\ &d((x_1 + S_W, y_1 + S_H), (x_2, y_2)), \\ &d((x_1 + S_W, y_1 - S_H), (x_2, y_2)), \\ &d((x_1 - S_W, y_1 + S_H), (x_2, y_2)), \\ &d((x_1 - S_W, y_1 - S_H), (x_2, y_2))\} \ . \end{aligned}$$

Here, S_W and S_H are the width and height of the smallest rectangle that contains network area S. This equation is illustrated by Figure 3.15. Assume two nodes, A and B, are both in a hexagon network and their positions are (x_1, y_1) and (x_2, y_2) respectively. The hexagon network area is contained by a rectangle with width S_W and height S_H. We replicate this rectangle in eight directions: right, left, bottom, up, right-bottom, right-up, left-bottom and left-up. Thus, node A has eight images each of which corresponds to a new position labeled in Figure 3.15. The toroidal distance between node A and B is defined to be the smallest euclidian distance among those from B to A and A's eight images.

With toroidal distance, nodes on the top of the network and nodes at the bottom cannot transmit concurrently. Neither can the left nodes and the right nodes. Another effect of toroidal distance is that all CSRs are constrained in the

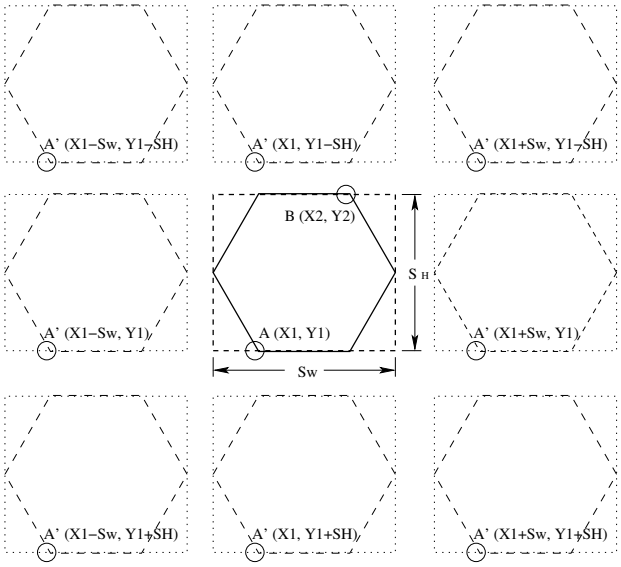


Figure 3.15: Toroidal distance

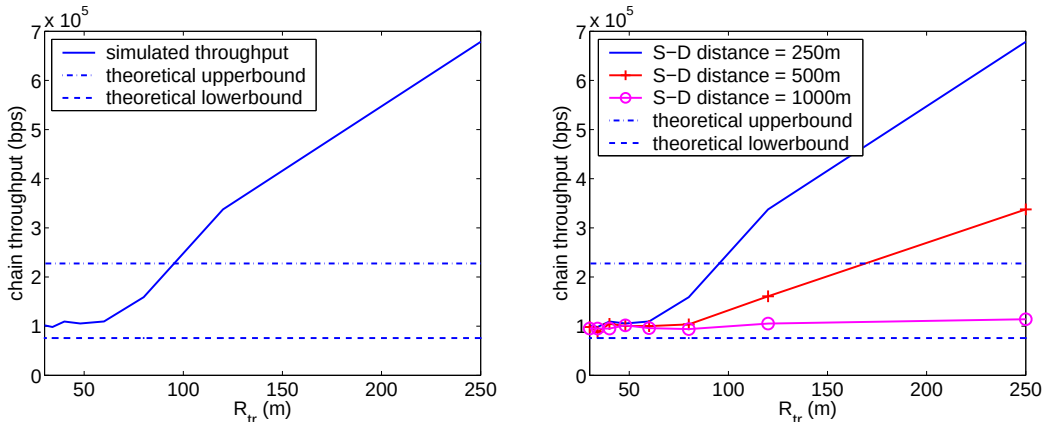
network area S. But, to eliminate the border effect, S should be greater than R_{cs} in 1D space, and both S_W and S_H should be greater than R_{cs} in 2D space, because, otherwise, S still could not cover the wrapped CSRs.

3.8 Experiment with Border Effect Elimination

We employ the two approaches proposed in Section 3.7 to the previous experiment, and demonstrate their effectiveness.

3.8.1 The Chain Traffic Flow

To allow a comparison, Figure 3.16(a) duplicates Figure 3.9. The dotted lines are lower and upper bounds plotted according to (3.55). The previous simulation was performed in a 250m network area, and the transmit range R_{tr} was varied from 250m to 100m. The network scale is small. Thus, the border effect dominated the experiment results. To eliminate the borders effect, we increase the S-D distance



(a) expriment with border effect

(b) expriment with border effect elimination

Figure 3.16: Eliminating the border effect in the chain traffic flow.

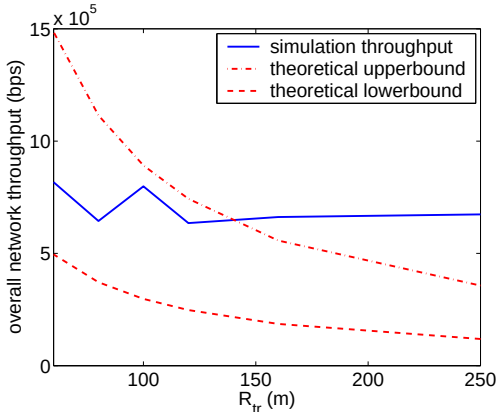
from 250m to 500m and 1000m so that the network area S is increased as well.

Figure 3.16(b) shows the simulation results. We observe that as the S-D distance increases to 1000m, the experiment results are strictly bounded by (3.55).

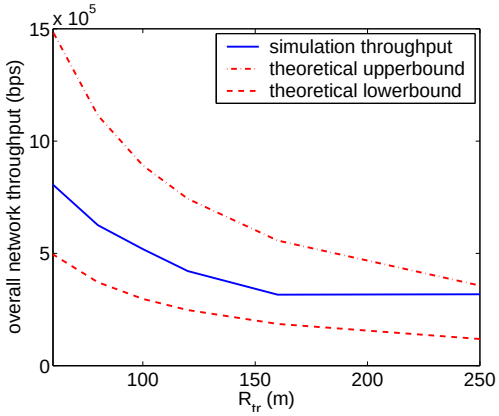
3.8.2 Multiple Parallel Traffic Flows

Figure 3.17(a) duplicates Figure 3.11 for comparison. The dotted lines are lower and upper bounds plotted according to (3.56). To eliminate the border effect, we increase the S-D distance of those traffic flows to $1000m$ and $1250m$ so that the network area S is increased to $1000m \times 1000m$ and $1250m \times 1000m$ respectively. In particular, we employ toroidal geometry when $S = 1000m \times 1000m$.

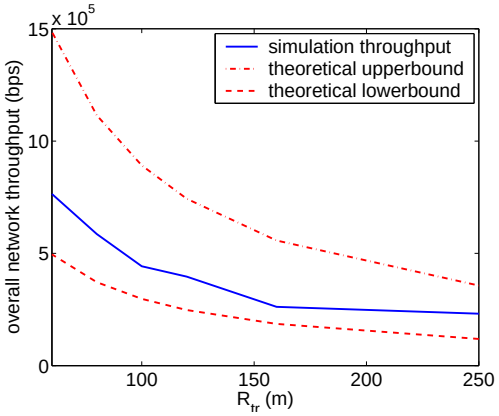
Figure 3.17(b) shows the simulation results of $S = 1000m \times 1000m$ with euclidian geometry. The throughput is almost bounded by (3.56). Figure 3.17(c) shows the simulation results of the same network area but with toroidal geometry. The throughput is bounded very well and its variations are close to the inverse of R_{tr}. Finally, we increase S to $1250m \times 1000m$ and the experimental results are strictly bounded by (3.56), as shown in Figure 3.17(d).



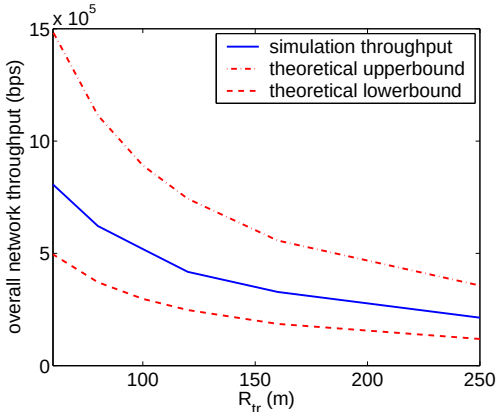
(a) S = 500m by 1000m, euclidian geometry



(b) S = 1000m by 1000m, euclidian geometry



(c) S = 1000m by 1000m, toroidal geometry



(d) S = 1250m by 1000m, euclidian geometry

Figure 3.17: Eliminate the border effect in the multiple parallel traffic flows.

Chapter 4

Optimized Transmit Power

Taking advantage of the insights gained by our analysis, we focus on research that restricts the strategies for improving spatial reuse to the MAC layer. The role of the MAC is to coordinate with the MAC's on other nodes to reserve the physical resources needed to transmit a given set of packets. Because we restrict ourselves to MAC only approaches, we cannot control the path taken by the packets being forwarded, but rather only the transmission at each individual hop. We will explore approaches that control routing strategies in Chapter 6 and 7.

Recall that the spatial usage metric Ω is defined to be $\Omega = \sum_{i=1}^{n_h} n_{ri} a_i t_i$, where n_h is the number of hops from source to destination, and n_{ri}, a_i, and t_i are respectively the number of retransmissions, the reserved area, and the transmission duration at the ith hop. The obvious method to improve spatial reuse at the MAC layer is reducing a_i so that the spatial usage of each individual transmission decreases. But, reducing a_i inappropriately could lead to the ongoing transmission being corrupted by co-channel interference and hence to a larger number of retransmissions, n_{ri}. Many MAC protocols do not balance this tradeoff very well. The reserved area a_i for an ongoing transmission is either larger or smaller than needed. In the former case, co-channel interference is reduced at the cost of inefficient spatial

usage. In the latter case, spatial reuse is improved at the cost of a higher likelihood of the received signal being corrupted by the tightly packed concurrent transmissions.

In our proposed MAC, we balance the reserved area a_i and the co-channel interference. Our goal is to find an appropriate size of a_i such that it is large enough to guarantee reliable reception but no larger. Employing this idea in a specific MAC protocol, the CSMA-based IEEE 802.11 DCF, we propose an enhanced MAC protocol to optimize a_i by adjusting the transmit power level. In CSMA-based MACs, optimizing a_i is in fact optimizing the size of the CSR. In identifying the optimal size of the CSR, we make two simplifications. First, we ignore the space reserved by the "virtual" carrier sensing (RTS/CTS handshake), because most of this area is covered by the CSR. Second, we only consider the co-channel interference caused by one hidden node. Our basic idea applies to aggregate co-channel interference, but complex. Moreover, we base our analysis on two key observations:

1. The exposed nodes that are not in the IR but silenced by carrier sensing should transmit for the sake of improving spatial reuse.

2. The hidden nodes that are in IR but not silenced by carrier sensing should not transmit for the sake of avoiding collisions.

Thus, balancing the spatial usage and co-channel interference is equivalent to balancing the exposed node and hidden node problems. By associating the potential number of exposed nodes and hidden nodes with the size of CSR and IR respectively, the tradeoff between minimizing the number of exposed nodes and the number of hidden nodes becomes the one between minimizing the size of CSR and IR at the same time. The optimal size of the CSR is identified to be large enough to cover IR but not more than that. Therefore, no hidden node would corrupt the ongoing transmission, and no redundant exposed node is reserved.

Using this optimization criteria, we derive the transmit power level corre-

sponding to the optimal CSR, and integrate this power control scheme, which we term *Optimized Transmit Power* (OTP), into the IEEE 802.11 MAC protocol. We evaluate the performance of OTP both analytically and experimentally. Our theoretical analysis shows that, for a random network, it conserves 10% spatial usage comparing to IEEE 802.11, which suggests a 10% potential network throughput improvement. However, simulation of a random topology shows marginal improvement less than 5%. A possible explanation is that most space conserved by OTP is not fully utilized by the traffic flows. This needs to be improved by using a more sophisticated routing protocol. We propose such approaches in Chapter 6 and 7.

4.1 Related Work

We first discuss the development history of the IEEE 802.11 MAC, and show that the exposed and hidden node problems are not well addressed. Thereafter, we discuss several existing power control schemes, and point out their weaknesses.

4.1.1 History of the IEEE 802.11 MAC

CSMA was a widely used scheme in packet radio networks [121]. In 1992, Kern [122] observed that the absence of a carrier does not mean a transmission will not interfere with a receiver, nor does the presence of a carrier mean that a transmission will interfere. He abandoned CSMA and instead adopted a RTS/CTS handshake to create Multiple Access with Collision Avoidance (MACA).

Later, Bharghavan [123] *et al.* improved MACA in several ways and proposed a new media access protocol called MACAW, which uses a RTS-CTS-DATA-ACK exchange and includes an exponential backoff algorithm. This work further specified the design basis: a simple physical model that ignores interference and the capture effect. In this model, any two nodes are either in-range or out-of-range of one another and all hidden nodes are assumed to be able to decode a CTS. In reality, this is not

true, nodes that are too far away from a receiver to decode a CTS can nevertheless transmit with enough power to lower the SINR enough to cause a collision [12].

Fullmer *et al.* [124] proposed Floor Acquisition Multiple Access (FAMA), which is the direct predecessor of the IEEE 802.11, by combining MACAW and CSMA. The goal was to eliminate the hidden node problem that remained in MACAW. Unfortunately, as we will show in detail, this was not very successful and both FAMA and the 802.11 MAC still have a hidden node problem. Furthermore, the re-addition of CSMA made the exposed node problem worse.

In retrospect, the exposed and hidden node problems are not addressed well by 802.11. A key point is that its development used a simple idealized physical model of wireless transmission consisting of two ranges: the TR and the CSR. It is becoming clear that this simple physical model is inadequate [12]. Xu [12] gives an analysis of the TR, the CSR and the IR, and makes it clear that all of these ranges are (partially) determined by the transmit power. To address the remaining hidden node problem, Xu proposed that a node replies with a CTS only when the received power is greater than a threshold, which is much larger than the receiving threshold, even if it is idle and receives the RTS successfully. Unfortunately, this approach can lead to otherwise avoidable network partitions.

4.1.2 Existing MAC Layer Power Control Schemes

MACs, like 802.11, that always transmit at full power, have the advantage of minimizing the IR but have the disadvantage of maximizing the CSR. In contrast, some MAC designs [85, 65] try to transmit with the minimum required power, generally in a attempt to conserve power, but also to enhance spatial reuse. The result is minimizing the CSR while maximizing the IR. Although spatial reuse is better, there is an increased potential for collisions due to hidden nodes.

A typical scheme sends an RTS with the maximum transmit power P_{max}.

Given the received power level P_r, the desired data packet transmit power P_t can then be calculated according to the following equation

$$P_t = c\mathcal{T}_{rx}\frac{P_{max}}{P_r} \ , \tag{4.1}$$

where $\mathcal{T}_{rx}$ is the receive power threshold for decoding a signal, and c is a constant slightly greater than 1 so that P_t is a little higher than the minimal desired power level. This scheme ignores co-channel interference from hidden nodes as well as other concurrent transmitters. In the simple model proposed in [123], this scheme works well because it is assumed that all of the hidden nodes receive and decode the CTS and thus remain silent. In reality, this scheme performs much worse than 802.11 because although the reduced transmit power reduces the exposed nodes, it greatly increases the hidden nodes. We give a detailed illustration of this in Section 4.2.1. It worth mentioning that, if channel fading is considered, the desired minimum power level should be higher.

Recently, some more sophisticated power control schemes have been proposed. Using a dedicated control channel or centralized infrastructure, they schedule concurrent transmissions based on receiver's interference tolerance level [81, 82]. (The details of those schemes are presented in Chapter 2.) However, a dedicated control channel requires extra transceiver devices at each node. This increases system cost and complexity. Unlike that work, we propose a simple and effective way of improving network spatial reuse without extra hardware cost and with trivial software modifications.

4.2 Design Rationale

OTP attempts to adjust transmit power level of each individual transmission so as to balance the spatial usage and co-channel interference, e.g. the size of the CSR

and IR. To explain OTP from a high level point of view, we introduce the design rationale, including the three range model and the optimized transmit power.

4.2.1 More About The Three-Range Physical Model

The limits of the 802.11 MAC and the need to consider each of the TR, CSR, and IR is becoming increasingly apparent. Extending the abstract concept of the three-range model introduced in Section 2.4, we present some insights and concrete analysis, which motivates the subsequent OTP scheme. Our studies are based on a combination of Free space and two-ray propagation model, e.g. the radio propagation conforms to the Friis free space model when the T-R distance is less than the reference distance d_{ref}, and the two-ray propagation model otherwise. The quantitative expression is

$$P_r = \begin{cases} \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L} & (d \leq d_{ref}) \\ \frac{P_t G_t G_r h_t^2 h_r^2}{d^4} & (d > d_{ref}) \end{cases}, \tag{4.2}$$

where G_t and G_r are the antenna gain of transmitter and receiver respectively, λ is the wavelength in meters, L is the system loss factor not related to propagation ($L \geq 1$), and h_t and h_r are the height of transmit antenna and receive antenna respectively.

Table 4.1 defines a variety of notations used in our derivations. In the case that values are given, they are based on the Lucent wireless card, which is used as the basis for the 802.11 model in ns2 [5, 6].

Note that our basic idea of improving spatial reuse is not restricted to either the radio propagation model or the exact value in Table 4.1. We choose them for two reasons. First, we need a radio propagation model and the parameters to demonstrate the idea. Second, using the same propagation model and parameters as those used by ns2 makes our subsequent simulations easier.

Notation	Meaning	Value
P_t	transmit power	
P_{opt}	transmit power defined by OTP	
P_{min}	the minimal required transmit power	
P_{max}	the maximum transmit power	0.2818w (24.5 dBm)
$\mathcal{T}_{rx}$	power level of receiving threshold	3.652×10^{-10}w (-64.4 dBm)
$\mathcal{T}_{cs}$	power level of carrier sensing threshold	1.559×10^{-11}w (-78.1 dBm)
$SINR_0$	SINR threshold of the capture effect	10 (10 dB)
G_t	Gain of the transmit antenna	1.0
G_r	Gain of the receive antenna	1.0
h_t	Height of the transmit antenna	1.5
h_r	Height of the receive antenna	1.5
d_{ref}	reference distance	87m
λ	wavelength	0.33m

Table 4.1: Notation and values where applicable.

The Transmit Range

With the specified propagation model, the radius of the transmit range, R_{tr}, is

$$R_{tr} = \begin{cases} \sqrt{\frac{P_t G_t G_r \lambda^2}{(4\pi)^2 \mathcal{T}_{rx}}} & R_{tr} < d_{ref} \\ \sqrt[4]{\frac{P_t G_t G_r h_t^2 h_r^2}{\mathcal{T}_{rx}}} & R_{tr} \geq d_{ref} \end{cases} . \tag{4.3}$$

We observe that for a particular transmission, in which the antennas and their location and heights are fixed, R_{tr} is solely a function of the transmit power P_t. The greater the transmit power, the larger the transmit range.

The Carrier Sense Range and Exposed Nodes

The radius of the CSR, R_{cs}, has the same derivation, except that $\mathcal{T}_{cs}$ replaces $\mathcal{T}_{rx}$. That is,

$$R_{cs} = \begin{cases} \sqrt{\frac{P_t G_t G_r \lambda^2}{(4\pi)^2 \mathcal{T}_{cs}}} & R_{cs} < d_{ref} \\ \sqrt[4]{\frac{P_t G_t G_r h_t^2 h_r^2}{\mathcal{T}_{cs}}} & R_{cs} \geq d_{ref} \end{cases} . \tag{4.4}$$

Again, for a particular transmission, R_{cs} is solely a function of P_t and when P_t increases, R_{cs} increases. Since all nodes inside R_{cs} must defer transmission, increasing P_t increases the region in which exposed nodes may occur and thus limits spatial reuse. This may degrade overall network throughput.

The Interference Range and Hidden Nodes

The derivation of the radius of the IR, R_{ir}, is a bit more complex because it must take into account not only the effects of the desired sender, but also the effects of a hidden node making an undesired transmission. In particular, to prevent a collision, the power level of the desired sender, P_r, and the power level of the interferer, P_{if}, at the receiver must satisfy the capture effect

$$\frac{P_r}{P_{if}} \geq SINR_0 \ . \tag{4.5}$$

We can use the propagation model to calculate P_r and P_{if}. We assume that the sender transmits with a (potentially variable) power of P_t from distance d and the worst possible interference such that the hidden node uses the maximum power, P_{max}, and transmits from distance R_{ir}. Using (4.5), we have

$$R_{ir} = \begin{cases} d\sqrt{\frac{P_{max}}{P_t} SINR_0} & (d < d_{ref}, R_{ir} < d_{ref}) \\ \sqrt[4]{\left(\frac{4\pi h_t h_r d}{\lambda}\right)^2 \frac{P_{max}}{P_t} SINR_0} & (d < d_{ref}, R_{ir} \geq d_{ref}) \\ d\sqrt[4]{\frac{P_{max}}{P_t} SINR_0} & (d \geq d_{ref}, R_{ir} \geq d_{ref}) \end{cases} . \tag{4.6}$$

Unlike the previous two ranges, R_{ir} is determined by both d and P_t. Increasing P_t will make the region in which hidden nodes may occur smaller, while increasing d will make it larger.

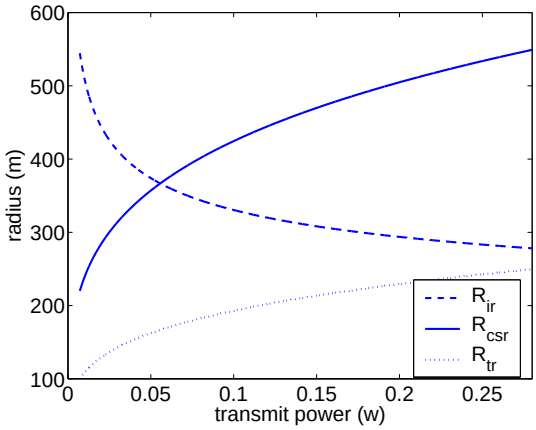


Figure 4.1: The relationship between the three ranges when $d = 100m$

4.2.2 The Relationship Between the Three Ranges

Figure 4.1 shows the quantitative relationships between the three ranges as P_t varies for a fixed d of 100m. To make the comparison more straightforward, we add d to R_{ir} so that the value of R_{ir} in the figure is the distance from the furthest hidden node to the sender instead of to the receiver.

We notice a dilemma in Figure 4.1. When P_t increases, R_{ir} and the area where hidden nodes can occur decreases, but R_{cs} and the area where exposed nodes can occur increases. However, to improve network throughput, we would like to reduce both the number of exposed nodes and the number of hidden nodes, or since we cannot know where these nodes are, the areas in which they may occur. This suggests that a good strategy would be to control P_t in such a way as to balance the hidden and exposed node problem.

Figure 4.1 gives us an intuition about what a good strategy might be. We observe that there is a crossing point between R_{ir} and R_{cs}, corresponding to a transmit power 0.056w (17.5 dBm). At this point, all potential hidden nodes will be within R_{cs} and thus will defer transmission and not cause collisions. If we decrease P_t then some hidden nodes may transmit and cause collisions. However, if we increase

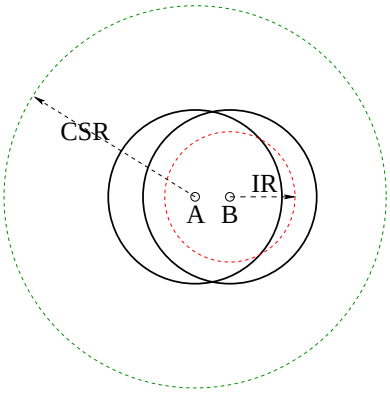


Figure 4.2: The three ranges with strong transmit power.

P_t we will not reduce the hidden node problem any further, but we will potentially create more exposed nodes.

4.2.3 Power Control Schemes

To reinforce the above observations, we begin by analyzing two existing power control schemes: the strong transmit power scheme and the weak transmit power scheme. We then propose our own scheme, the *Optimized Transmit Power* (OTP) scheme. In general, we argue that inappropriate transmit power choices can degrade network throughput.

Strong Transmit Power

The IEEE 802.11 DCF [1] is the prototypical MAC using strong transmit power. All nodes transmit at P_{max} no matter how close the sender and receiver are. To illustrate, consider the case where d is 100m and P_{max} is 0.2818w (24.5 dBm). Then according to (4.3), (4.4) and (4.6), R_{tr}, R_{cs} and R_{ir} are 250m, 550m, and 170m respectively, as shown in Figure 4.2. We show TR for both node A, the transmitter, and node B, the receiver, since this shows both where the RTS and data can be received and where the CTS and ACK can be received. Obviously, IR_B is enclosed

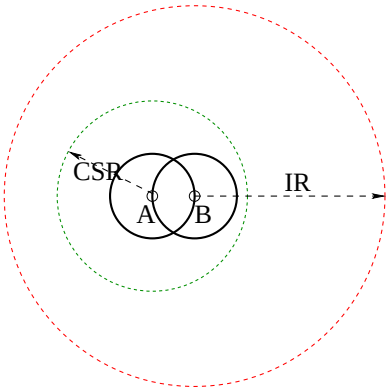


Figure 4.3: The three ranges with weak transmit power.

by not only CSR_A but also TR_B, so all possible hidden nodes defer transmission both as a result of decoding the virtual carrier information in the RTS/CTS as well as a result of sensing the physical carrier. On the other hand, CSR_A covers significantly more area than just IR_B, which means that there is significant potential for exposed nodes. This limits spatial reuse.

Note that Figure 4.2 is plotted according to the parameters of the Lucent wireless card, where the ratio, γ, of R_{cs} and R_{tr} is greater than 2. Thus, both TR_A and TR_B are inside CSR_A no matter what P_t and d are. When $\gamma < 2$, TR_B, the CTS range, may reserve area outside CSR_A, which generates extra exposed nodes. But, as long as $\gamma > 1$, the area of $(TR_B - CSR_A)$ is neglectable.

Weak Transmit Power

Many power control schemes use (4.1) to calculate the desired transmit power, which leads to the weak transmit power scenario. We call this the Minimal Transmit Power (MTP) scheme. If d is 100m, then the desired transmit power is $P_t = 0.0072w$ according to (4.1). Therefore, R_{tr}, R_{ir} and R_{cs} are 100m, 450m and 220m respectively, as shown in Figure 4.3. Now, IR_B extends beyond TR_A, TR_B, and CSR_A. This means there are potential hidden nodes that do not acquire the virtual carrier sense

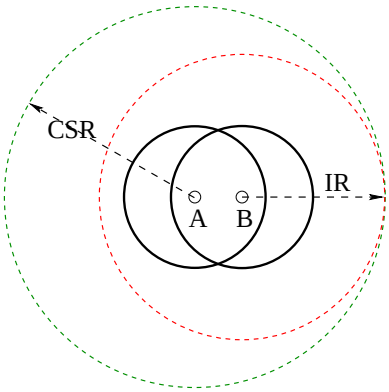


Figure 4.4: The three ranges with an optimized transmission power

from either the RTS or CTS and which also cannot sense the physical carrier. If these hidden nodes transmit, they will create collisions at B, again degrading network throughput.

Note that Figure 4.3 is plotted according to the parameters of Lucent wireless card, where the ratio, γ, of R_{cs} and R_{tr} is greater than 2. Thus, both TR_A and TR_B are inside CSR_A no matter what P_t and d are. When $\gamma < 2$, TR_B, the CTS range, may help CSR_A reserve some area inside IR_B and eliminate some hidden nodes. But, as long as $\gamma > 1$, the area of $(TR_B - CSR_A)$ is neglectable.

Optimized Transmit Power

We identify an optimized transmit power to optimize the size of the CSR so that it balances the hidden and exposed node problems. The idea should be obvious by now; we choose the optimal transmit power level P_{otp} so that CSR_A will exactly cover IR_B. This means all potential hidden nodes will have their transmissions suppressed due to CSMA, but CSR_A will be no larger than needed to achieve this goal, thus avoiding the creation of unneeded additional exposed nodes. This scenario is shown in Figure 4.4.

The desired condition satisfies the following equation

$$R_{cs} = d + R_{ir} \ . \tag{4.7}$$

Substituting (4.4) and (4.6) into (4.7), we obtain a function for calculating the optimized transmit power P_{otp}, e.g.

$$P_{otp} = \begin{cases} \left(\frac{d+\sqrt{d^2+4AC}}{2A}\right)^4 & d < 200m \\ P_{max} & d \geq 200m \end{cases} , \tag{4.8}$$

where

$$A = \sqrt[4]{\frac{G_t G_r h_t^2 h_r^2}{\mathcal{T}_{cs}}} \tag{4.9}$$

and

$$C = \begin{cases} \sqrt[4]{\left(\frac{(4\pi)h_t h_r d}{\lambda}\right)^2 P_{max} SINR_0} & d < d_{ref} \\ d\sqrt[4]{P_{max} SINR_0} & d \geq d_{ref} \end{cases} . \tag{4.10}$$

In (4.8), we set P_{otp} to P_{max} when d exceeds 200m, because in the previous derivation ((4.6)) we assume that the maximal transmit power is P_{max}. Thus no node can be allowed to transmit at power level greater than P_{max}. Therefore, when d is greater than 200m, our power control scheme is no different than 802.11 and for these distances will show no advantage compared to 802.11.

Note that P_{otp} is derived such that it is strong enough to counter single co-channel interferer. Deriving the desired transmit power of countering aggregate co-channel interference is possible. Using the worst case co-channel interference model illustrated by Figure 3.2, the desired transmit power level is the one that guarantees the lowest possible SINR higher than $SINR_0$. Because there are not many interesting issues in this derivation and aggregate interference from multiple hidden nodes may be less likely due to the fact of carrier sensing, we do not present

in this thesis. As an alternative, we can simply multiply (4.8) by a constant which is greater than 1 or add a constant which is greater than 0.

4.2.4 OTP and Rate Adaptation

By controlling transmit power, OTP balances the size of the CSR and IR such that CSR is large enough to cover IR and not larger. This goal can also be achieved by using rate adaptation.

For a given transmit power and a given sender-receiver distance, increasing transmission rate increases the SINR threshold at the receiver and, therefore, increases the IR. Vice versa, decreasing transmission rate decreases the SINR threshold and, therefore, reduces the IR. Because the CSR is a function of transmit power and carrier sensing threshold, which does not change with transmission rate, the CSR does not change with transmission rate. The sender can choose the highest possible rate that satisfies (4.7), namely the *Optimized Transmission Rate* (OTR).

An interesting question is which strategy brings about more significant improvement to network performance, OTP or OTR? Or, is there an optimized combination of the two strategies that maximizes network performance? According to Section 3.4, such optimization problems depend significantly on the specific traffic pattern and network topology. Specifically, OTP is more efficient for spatially limited networks and OTR is more efficient for spatially unlimited networks.

Finally, an observation is that, for a given sender-receiver distance, transmission rate can be raised further if the power level is allowed to increase. In doing this, high transmission rate is achieved at the cost of a larger CSR of each individual transmission. Again, as we point out in Section 3.4, the combined effect is determined significantly by specific traffic pattern and research into achieving a good balance is left for future work.

4.3 The OTP Protocol

We first present the details of the protocol and then discuss some feasibility issues.

4.3.1 Details of the OTP Protocol

We use the 802.11 DCF MAC to demonstrate how the idea above could be integrated into a CSMA-based MAC and propose the *Optimized Transmit Power* (OTP) protocol. Recall that all of the following transmissions must wait for an idle channel. The protocol proceeds as follows:

1 The sender sends an RTS using P_{max}.

2 The receiver estimates the receive signal power P_r, piggyback P_r in a CTS and replies with the CTS using P_{max}.

3 The sender calculates the distance, d, from the receiver by using the received power level, P_r, and P_{max}. It then calculates the optimized transmit power level P_{otp} according to (4.8). Finally, the sender transmits the data packet using P_{otp}.

4 Upon receiving the data packet, the receiver replies with an ACK using P_{otp}. It calculates P_{otp} in the same manner as the sender.

4.3.2 OTP and Distance Estimation

Now, we discuss the feasibility of steps 2 and 3 in the protocol presented above. The essential issue is how we obtain the distance, d, between the sender and the receiver. A potential solution would be using the positioning techniques presented in Section 2.10.

If we do not use positioning techniques, we need to estimate the distance by exchanging control packets and measuring their transmit and receive power levels,

as specified in the above protocol. Thus, the issue that needs to be addressed is how to measure the receiving signal power level, P_r, at the receiver, because the total power at the receiver, Y, is a combination of P_r, the noise N_0 and the interference I, e.g. $Y = P_r + N_0 + I$. Section 2.8 presents multiple sophisticated SINR estimation techniques that can be used to measure the signal power. Once the transmit and receive power are known, we can estimate the sender-receiver distance and calculate P_{otp}. The limitation is that we need to know the specific large-propagation path loss model.

Additionally, Section 2.9 presents two models of approximating the signal power. Using the lower bound in (2.13) or (2.14), we may overestimate P_{otp}. Using the upper bound in (2.13) or (2.14), we may underestimate P_{otp}. Chapter 5 shows that OTP is overly conservative. Therefore, we can choose to either overestimate or underestimate P_{otp}, depending on whether we want OTP to be more conservative or less conservative.

4.4 Performance Analysis

Now, we analyze the performance of OTP. Specifically, we compare the potential network throughput of using OTP and that of using the standard IEEE 802.11 in which all nodes transmit with P_{max}. Chapter 3 shows that the the spatial usage metric, Ω, is an indication of potential overall network throughput. Hence, our strategy is to find out how much space can be conserved as an indication of potential network throughput improvement.

The spatial usage metric Ω is

$$\Omega = \sum_{i=1}^{n_h} a_i n_{ri} t_i \ , \tag{4.11}$$

where n_h is the number of hops, and a_i, n_{ri} and t_i are the size of CSR, the number of

retransmissions and the transmit duration for the ith hop. Therefore, Ω is the total amount of space-time consumed by forwarding a packet from source to destination.

t_i: We assume that the transmit duration t_i at each hop is identical in both OTP and 802.11.

n_{ri}: We also assume that n_{ri} is 1 in both OTP and 802.11. This assumption is reasonable because both OTP and 802.11 attempt to eliminate the most severe interference that comes from hidden nodes.

n_h: Given the S-D locations, n_h is identical for both OTP and 802.11, because they are simply MAC layer protocols and they use the same network layer routing protocol to discovery pathes.

a_i: In 802.11, a_i is πR_{max}^2 for every single transmission, where R_{max} is the carrier sense range of P_{max}. In OTP, a_i is πR_{otp}^2, where R_{otp}^2 is the carrier sense range of P_{otp}. It can be derived from (4.4) and (4.8), e.g.

$$R_{otp} = \begin{cases} \frac{d+\sqrt{d^2+4AC}}{2} & d < 200m \\ R_{max} & d \geq 200m \end{cases}, \tag{4.12}$$

where A and C are specified in (4.8).

Therefore, the spatial usage of OTP and 802.11 are respectively

$$\begin{cases} \Omega_{otp} = \sum_{i=1}^{n_h} t\pi R_{otp}^2 = t\pi \sum_{i=1}^{n_h} R_{otp}^2 \\ \\ \Omega_{802.11} = \sum_{i=1}^{n_h} t\pi R_{max}^2 = t\pi n_h R_{max}^2 \end{cases}. \tag{4.13}$$

Further, we define E to be the ratio of Ω_{otp} and $\Omega_{802.11}$. That is,

$$E = \frac{\Omega_{otp}}{\Omega_{802.11}} = \frac{1}{n_h R_{max}^2} \left(\sum_{i=1}^{n_h} R_{otp}^2 \right) = \frac{1}{R_{max}^2} \left(\frac{1}{n_h} \sum_{i=1}^{n_h} R_{otp}^2 \right) = \frac{\overline{R_{otp}^2}}{R_{max}^2} \ . \qquad (4.14)$$

Thus, the critical issue is finding $\overline{R_{otp}^2}$. We know that R_{otp} is a function of the T-R distance d, and the physical layer T-R distance is also the network layer hop distance. The Monte Carlo simulations presented in Appendix B suggest that, for networks in which nodes are uniformly distributed, the hop-distance distribution can be approximated as an exponential distribution

$$p(d) = \frac{5}{R_{tr}} e^{-\frac{5}{R_{tr}}(R_{tr}-d)} \ , \qquad (4.15)$$

where R_{tr} is the transmit range of the power level at which the shortest path is discovered. We consider the case that R_{tr} is 250m which corresponds to the power level P_{max}.

Note that OTP is triggered only when d is less than a threshold. According to the parameters specified in Table 4.1, this threshold is 200m. Moreover, d varies from 0 to R_{tr}. With (4.12), (4.14), and (4.15), we have

$$\begin{aligned} E &= \tfrac{1}{R_{max}^2} \int_0^{R_{tr}} p(x) R_{otp}^2 dx \\ &= \tfrac{1}{R_{max}^2} \left(\int_0^{200} p(x) R_{otp}^2 dx + \int_{200}^{R_{tr}} p(x) R_{max}^2 dx \right) \qquad (4.16) \\ &\approx 0.9 \ . \end{aligned}$$

The value of this integration was obtained by Monte Carlo simulations. It shows that, to forward a packet from source to destination, the average spatial usage of OTP is approximately 90% of 802.11. Therefore, the potential network throughput improvement is approximately 10%.

4.5 Simulation Results

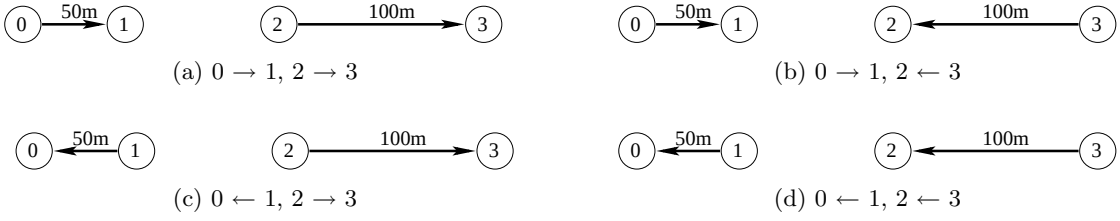
To compare OTP to 802.11 and MTP, we implemented each of them in ns2 [5, 6]. As Jung [67] did, we assume that MTP transmits the RTS/CTS with P_{max} and DATA/ACK with P_{min}, which is calculated by (4.1). The constant c in (4.1) is set to 1. The implementations of MTP and OTP are essentially the same. The MAC obtains the received power levels of the RTS/CTS from the physical layer, calculates the estimated distance and desired transmit power according to (4.1) or (4.8), passes it to the physical layer, and the physical layer transmits the DATA/ACK with the desired power. In either case, broadcast packets, including those used for routings, are transmitted with P_{max}.

We then performed simulations using three different network topologies: two traffic flows, chains, and random. We choose the topology of two traffic flows because it is the simplest one for studying "inter-flow" spatial reuse, the chain topology because it is the typical one for studying "intra-flow" spatial reuse, and the random topology because it represent more general cases.

4.5.1 Two Traffic Flows

Figure 4.5 shows four combinations of flow directions. The two traffic flows are Tr_{01} with $d = 50m$ and Tr_{23} with $d = 100m$. Both senders inject CBR traffic at the full channel capacity. The power levels and R_{tr}, R_{ir}, and R_{cs} for each flow are listed in Table 4.2. We varied the distance between node 1 and 2, d_{12}, from 100m to 600m and measured the network throughput for each scheme.

We simulated all four combinations of flow directions. The results are shown in Figure 4.6. The X-axis is d_{12} and the Y-axis the throughput normalized by the channel capacity. Because the explanation of these results are similar, we only explain in detail the one of Figure 4.6(a) which corresponds to the topology of Figure 4.5(a). We discuss each scheme in turn.



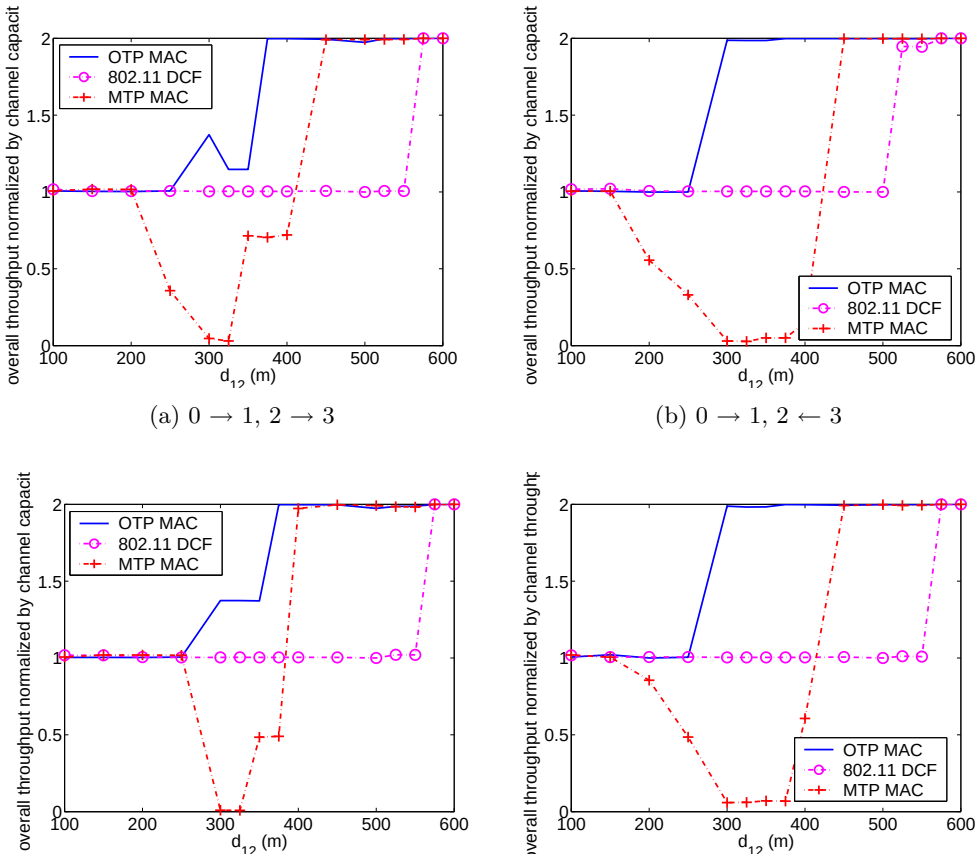
(a) $0 \rightarrow 1$, $2 \rightarrow 3$ (b) $0 \rightarrow 1$, $2 \leftarrow 3$

(c) $0 \leftarrow 1$, $2 \rightarrow 3$ (d) $0 \leftarrow 1$, $2 \leftarrow 3$

Figure 4.5: Two traffic flows with four combinations of flow directions.

	OTP MAC	IEEE 802.11	MTP MAC
d	50m	50m	50m
P_t	0.0188w (12.7 dBm)	0.2818w (24.5 dBm)	4.5×10^{-4}w (-3.5 dBm)
R_{tr}	127m	250m	50m
R_{ir}	230m	89m	445m
R_{cs}	280m	550m	109m
d	100m	100m	100m
P_t	0.0557w (17.5 dBm)	0.2818w (24.5 dBm)	0.0072w (8.6 dBm)
R_{tr}	166m	250m	100m
R_{ir}	266m	178m	445m
R_{cs}	366m	550m	220m

Table 4.2: Radiuses under different power control schemes.



(a) $0 \rightarrow 1,\ 2 \rightarrow 3$ (b) $0 \rightarrow 1,\ 2 \leftarrow 3$

(c) $0 \leftarrow 1,\ 2 \rightarrow 3$ (d) $0 \leftarrow 1,\ 2 \leftarrow 3$

Figure 4.6: Throughput for two traffic flows.

IEEE 802.11

802.11 only achieves maximum performance when d_{12} exceeds 550m, which is to be expected since that is R_{cs}. Notice that since R_{ir} is only 89m, node 0 and node 2 could have begun simultaneous transmission at that distance if it were not for the excessively large CSR.

MTP MAC

The primary drawback of MTP is that the IR greatly exceeds the CSR when sending data and therefore data has a significant chance of being corrupted at the receiver. To understand these results in detail, we analyze several critical values of d_{12} focusing on the effects of the IR.

$d_{12} <= 200m$: Both node 0 and 2 can receive each others RTS. Thus they alternate transmission and the two flows split the channel capacity.

$d_{12} = 300m$: Node 0 can neither receive the RTS/CTS from Tr_{23} nor detect the carrier from node 2. The same is true of node 2 with respect to Tr_{01}. Hence, both node 0 and 2 may commence transmission simultaneously. However, both node 2 and 3 are in IR_1, so the RTS/CTS transmissions of Tr_{23} cause collisions at node 1. Similarly, node 1 is in IR_3 and IR_2. Thus a CTS from node 1 can corrupt data at node 3 or a CTS or ACK at node 2. As a result, network throughput is severely degraded. Notice that because data and ACKs are transmitted at low power, it is the RTS/CTS exchange that results in corruption.

$d_{12} = 350m$: Again, node 0 and 2 can transmit simultaneously. Node 2 is in IR_1 and so its RTS corrupts data at node 1. Thus, Tr_{01} makes almost no contribution to the overall throughput, which is an unfairness phenomena that occurs occasionally in 802.11 DCF based networks. Furthermore, Tr_{01} may corrupt ACKs at node 2, even though it does not corrupt data at node 3. Hence, Tr_{23} is partially degraded, and the overall throughput is slightly worse than the channel capacity.

$d_{12} = 450m$: Again, both flows can transmit simultaneously. However, now all the nodes are outside each others IR and no collisions occur. Because data and ACKs are transmitted with lower power than in 802.11, they do not cause the undue carrier sensing we saw in 802.11. Finally, although the RTS and CTS are transmitted at full power, they are brief and infrequent and so do not cause undue carrier sensing either. Thus, spatial reuse is fully achieved.

OTP MAC

For the OTP scheme, the CSR is used to eliminate hidden nodes. Again, we look at several critical values of d_{12} but this time, since there are no hidden nodes, the issues focus on the CSR.

$d_{12} = 250m$: Node 0 and 1 are in CSR_2. Node 2 can receive the CTS from node 1, even though it is out of CSR_0. Therefore Tr_{01} and Tr_{23} utilize the channel alternatively, and there is no spatial reuse.

$d_{12} = 300m$: Node 2 and 3 are out of CSR_0, while node 0 and 1 are both in CSR_2. Further, node 2 can no longer receive the CTS from node 1 and it is outside IR_1. Therefore, node 2 can transmit at the same time as node 0 without corrupting the signal at node 1. But node 0 has to defer while node 2 is transmitting. Thus depending on the timing sometimes both node 0 and 2 can transmit and sometimes only node 2 can. Partial spatial reuse is achieved, and throughput exceeds the channel capacity, but does not achieve twice the capacity.

$d_{12} = 325m$: Network throughput drops at this point because of 802.11's exponential backoff scheme. Node 1 is in CSR_2, while node 0 is out of it. Node 0 may send an RTS while node 2 is transmitting. Now node 1 cannot reply with a CTS due to carrier sense, which causes node 0 to exponentially backoff and then retransmit the RTS. This results in an increasingly small chance that node 0 will gain access to the channel. As a result, Tr_{01} has a slight contribution to overall

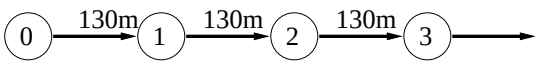


Figure 4.7: The chain topology

network throughput, but Tr_{23} can fully utilize channel capacity.

$d_{12} = 375m$: Starting at this point, both flows are outside of the others CSR. Spatial reuse is fully achieved, and the network throughput reaches two times channel capacity significantly sooner than either of the other two schemes.

4.5.2 Chain Topologies

Our second study focuses on chain topologies, as shown in Figure 4.7. We used a minimum hop-count routing protocol, DSR [35].

It is obvious that the closer the spacing, the better the performance of OTP. But if the distance is less than $0.5R_{tr}$ which is 125m in the simulation setup, DSR would skip the intermediate node and go directly to the next one. Therefore we separate the nodes by 130m, in which case R_{cs} is 550m for 802.11, 430m for OTP, and 286m for MTP.

The source node 0 attempts to inject packets at the full channel capacity. We vary the number of hops from 2 to 9. When the hop count exceeds 6, the theoretical throughput upper bound for 802.11 is between $\frac{1}{5}$ and $\frac{1}{6}$ of the channel capacity if there is perfect scheduling so that transmissions that are 5 hops away transmit simultaneously [125]. For example, Tr_{01} and Tr_{56} transmit at the same time, and Tr_{12} and Tr_{67} transmit at the same time, etc.

The results are shown in Figure 4.8, where the X-axis is the number of hops and the Y-axis is the throughput normalized by the channel capacity. We show the results for the theoretical upper bound as well as the three schemes. We observe that as the hop count increases from 2 to 5, all the curves decrease sharply. Although MTP can achieve spatial reuse starting from hop count 4, this gain is nullified by

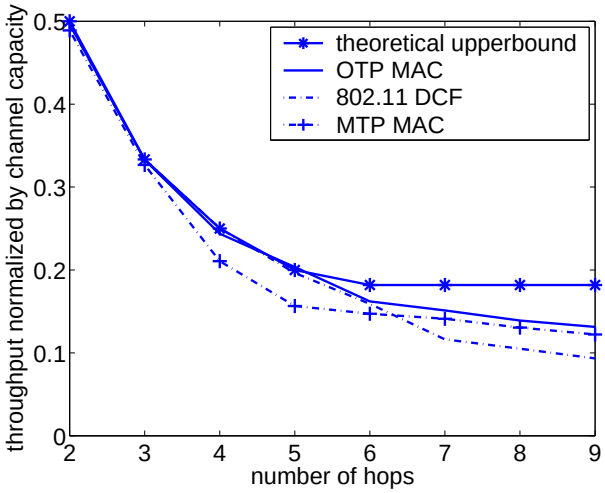


Figure 4.8: Throughput for the chain topology.

interference from hidden nodes, much as we saw in the previous section. OTP's throughput begins to flatten out at 6 hops, which is when a second node in the chain can begin transmitting at the same time as node 0, while 802.11, with it's greater CSR, does not flatten out until 7 hops. Thus, although the advantages are modest, the OTP MAC out performs both of the other schemes and for the reasons we expected.

4.5.3 Random Topologies

To test more general topologies, we placed nodes randomly in a 1000m by 1000m area with four pairs of randomly chosen nodes acting as source and destination. Again, the ns-2 implementation of DSR is used as the routing algorithm and each source attempts to transmit at the full channel capacity. Because we are interested primarily in how much spatial reuse is achieved, nodes are not mobile. A key claim is that OTP increases spatial reuse and we investigated this claim by varying the number of nodes from 10 to 50. We generated 300 random configurations for each density. Our results are the average over these configurations.

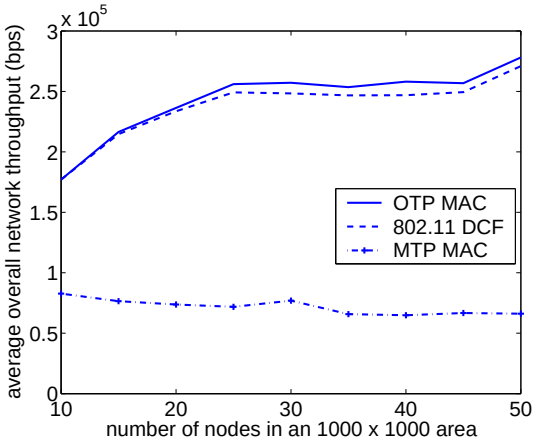


Figure 4.9: Throughput for random network topologies.

Our results are shown in Figure 4.9, where the X-axis is the number of nodes and the Y-axis is the throughput in bps. In general, OTP and 802.11 give similar results, while MTP performs significantly worse due to its unsolved hidden node problem.

OTP did not improve performance substantially. Preliminary analysis indicates two explanations. First, using the shortest-path routing protocol, many hops are close to or greater than the 200m limit at which OTP gives no advantage over 802.11. This limits the spatial reuse from OTP. It also suggests future work in routing techniques that use short hop distance routes. Second, our theoretical analysis shows 10% potential spatial conservation, but the experimental results shows 5% improvement to network throughput. A possible explanation is that some of the space conserved by OTP is not utilized by the traffic flows. A simple example is that there are not transmissions in the conserved space.

Nevertheless, we are able to observe that the advantage of OTP over 802.11 does increase with increasing density. Thus although subtle we do get the increased spatial reuse we would expect.

Chapter 5

Enhancing the OTP Scheme

Previously, we presented a MAC layer power control scheme, OTP, to improve network space utilization by balancing the size of the CSR and the IR for each individual transmission. In practice, this scheme is overly conservative in two ways:

- First, when possible, OTP always transmits data packets at the power level that guarantees successful packet reception even when a receiver experiences the worst possible interference, e.g. an interferer transmitting at the maximum power residing just outside the CSR. Since such a worst case does not occur much of the time, in practice, the transmit power can be reduced further.

- Second, OTP exchanges RTS/CTS packets at the maximum power, which creates the maximum possible interference to ongoing transmissions. This makes it easy to estimate the sender-receiver distance, given the assumption that we ignore fast fading and that we know the exact path loss model. But it is not necessary to transmit the control packets at the maximum power. Under the same assumptions, the distance can be derived as long as the transmit and receive power levels are known. Therefore, both the RTS and CTS can be sent at a lower power level just as the ACK and data packets are.

Motivated by these two observations, we propose an *Enhanced OTP* (EOTP) protocol, which dynamically adjusts transmit power according to the actual interference level at the receiver. Some existing MAC power control schemes employ this idea and tune transmit power according to the previous instantaneous interference at the receiver [81, 82, 84, 115]. Because the transmission duration of a data packet is usually longer than that of a control packet, the instantaneous interference measured during an RTS/CTS exchange may result into a low power level for the subsequent data transmission. To solve this problem, the previously proposed MAC power control schemes use busy tones on a control channel to protect data packets from being corrupted. In single channel networks, there is no dedicated control channel and it would be risky to adjust transmit power according to the instantaneous interference. Therefore, the sender needs to observe the interference at the receiver for a relatively long period and adjust transmit power accordingly. In addition, we will argue that exchanging control packets at the maximum power level brings about no extra benefit, and thus we reduce their transmit power to avoid generating unneeded interference.

In fact, the motivation for enhancing OTP come from our subsequent investigation of routing strategies in Chapter 6. We observed that choosing short hops did not improve network throughput as we expected. Studying the simulation results showed that OTP is overly conservative in many situations. This leads to the development of EOTP.

5.1 Related Work

A number of MAC layer power control schemes have been proposed to adjust transmit power according to the actual interference level at the receiver. Most of them use multiple channels [81, 82, 83, 84, 115]. Wu *et al.* [81] proposed that the sender transmits at the minimum required power estimated during an RTS/CTS exchange

and that the receiver sends out a busy tone at the maximum power in the control channel to protect an ongoing transmission. Similarly, Monks *et al.* [82] leveraged a request-power-to-send (RPTS)/acceptable-power-to-send (APTS) exchange to determine the minimum transmit power that will result in a possible successful packet reception. To protect an ongoing transmission, a receiver sends busy tone pulses on a control channel. The signal strength of the busy tone indicates the receiver's tolerance to additional noise. A potential transmitter decides the upperbound of its transmit power by sensing the busy tones. Lin *et al.* [83] enhanced the IEEE 802.11 standard by improving the handshake mechanism and adding a control channel to notify the neighbors around a receiver about the noise tolerance. Thus, the neighbors can adjust their transmit power to avoid packet collision at the receiver. In [115], the RTS/CTS are exchanged on the control channel at the maximum power. A receiver uses the CTS to notify its neighbors of the additional noise power that can be added without impacting data reception.

Due to the utilization of multiple channels, the schemes above have a more complicated physical layer implementation than those that use a single channel. We prefer a MAC power control scheme using a single channel for the sake of hardware and physical layer simplicity. An added benefit is that the scheme could share the same physical layer standard as the IEEE 802.11.

Few single channel based power control schemes have been proposed [13, 85, 86, 87, 88, 7]. In single channel wireless networks, it is a challenge to guarantee successful packet reception by sending packets at a power level adjusted according to past interference information, especially when the interference information is instantaneous and measured over a short interval. In Chapter 4 [13], we propose that the sender always assumes the highest interference at the receiver and tunes the transmit power such that the CSR is just large enough to cover the IR. As we stated above, our OTP is overly conservative in practice. Inspired by cellular

CDMA power control schemes, Agarwal *et al.* [85] proposed to ratchet up (or down) transmit power by observing the exchange and loss of messages. Even though this is a simple strategy, its power tuning scheme is not justified. For example, the authors simply assume that retransmission is an indication of packet loss and high interference, but they ignore the fact that retransmission may be caused by a receiver that is not able to reply with either a CTS or an ACK. Poon *et al.* [86] proposed that, after special handling of control packets exchange, the data packets are transmitted at the minimum required power level. Although this is a simple strategy, transmitting data packets at the minimum required power level makes the transmitted signal vulnerable. Yu *et al.* [87] proposed *Power-Stepped Protocol* (PSP) to enhance spatial utilization in clustered mobile ad hoc networks. In PSP, each node can operate at a different power level but not more than one level higher or not less than one level lower than that of any of its neighbors. The algorithm specifies that a node increases its power when the number of neighbors is less than a predefined threshold and decreases its power when the number of neighbors is greater than a predefined threshold. Thus, PSP is sensitive to node density and is suitable in a clustered topology. The drawback of PSP is that it uses periodic hello messages to exchange neighbor set information and, therefore, may result in substantial overhead. In the framework of the IEEE 802.11 standard, Zhang *et al.* [88] studied the correlation between the necessary transmit power of RTS, CTS, DATA and ACK packets. Their essential idea is to use virtual carrier sense to suppress the transmission of the potential interferers and, therefore, guarantee the next packet reception. For example, the transmission of CTS clears a floor around the receiver so that the upcoming data packets won't be corrupted. Virtual carrier sense works well if all the neighbors in the transmit range decode the transmitted signal correctly and, therefore, keep silent for the rest of the packet exchange. However, it is impossible for all the neighbors to correctly decode the transmitted signal due to kinds of reasons such as

Notation	Meaning
P_{min}	the minimal transmit power.
P_{max}	the maximal transmit power allowed in a network.
P_{opt}	the transmit power used by OTP scheme.
P_{eotp}	the transmit power used by EOTP scheme.
d	distance between the sender and the receiver.
d_{max}	the transmit range of P_{max}.

Table 5.1: List of notation.

collisions, etc. Muqattash *et al.* [7] proposed a rather complicate power control MAC protocol, called POWMAC, which uses an access window (AW) to schedule several concurrent data packet transmissions by using a series of RTS/CTS exchange. In POWMAC, the transmit power is set such that the received signal power at the receiver is high enough to not only overcome current interference but also allow a certain amount of additions to overcome upcoming concurrent transmissions that might be scheduled in the vicinity. Therefore, POWMAC reserves a larger than needed CSR for an individual transmission. Other transmissions within this larger CSR are scheduled using a series of RTS/CTS exchanges within the AW interval. A potential limitation of POWMAC is the AW synchronization among neighboring nodes, especially for those that are two hops away, e.g., the nodes that are inside the CSR but out of TR. These two-hop-away nodes should be able to transmit if the transmit power is lower than that specified in POWMAC.

5.2 Design Rationale

The basic idea of EOTP is for a sender to observe the interference level at a receiver for a relatively long duration and, thereafter, to tune transmit power accordingly. To explain EOTP from a high level point of view, we introduce the design rationales behind it. To facilitate our description, we use the notation listed in Table 5.1, where the definitions of P_{min}, P_{max}, and P_{otp} are the same as in the previous Chapters.

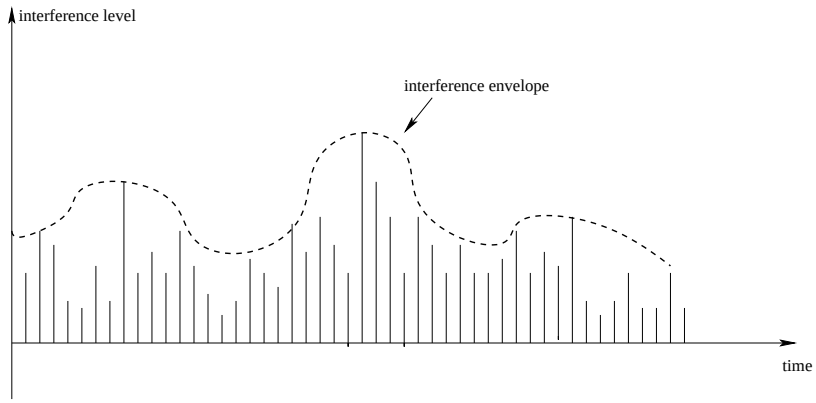


Figure 5.1: Illustration of the interference envelope.

5.2.1 Estimating the Interference at the Receiver

Because OTP is based on preventing the worst possible case, the details of the level of interference at the receiver are unimportant. This is not the case for EOTP and so it is important to consider conditions at the receiver. The essential issue is whether or not the SINR ever falls below the threshold. Thus, we are concerned about the highest level of interference, or in the envelope defined by peaks in the interference level, as shown in Figure 5.1. It is not simply enough to track the interference envelope; there are several subtle issues to consider.

The first issue is when does the envelope matter? The most obvious point is that the level of interference only matters when a transmitter is actually trying to communicate with the receiver. Further, because of CSMA, a transmission will actually suppress many of the closest possible interferers. This implies that when we change the power level, the exact set of possible interferes may well change. Thus, ideally, we should track the envelope on a transmitter-receiver pair basis and at the actual power level to be used.

The second issue is what time scale matters? We want to make it unlikely that a collision occurs, but if it allowed a significantly lower power level, we might tolerate a rare collision. Typical data packet transmissions are significantly longer

than control packet transmissions. Thus if we observe the envelope only during control packet transmissions we are likely to underestimate the maximum interference that might occur during a data packet transmission. Likewise, if we track the maximum interference for a very long time, it is likely that we will overestimate the interference that will typically occur. Thus, we argue that we should observe the envelope over some bounded and relatively small number of data packet intervals.

5.2.2 Two Approaches to Obtain Interference Information

There are two obvious ways for the sender to obtain interference information and, therefore, two ways of varying the transmit power. Both of the approaches observe the interference for a relatively long duration before tuning transmit power.

One approach, which we call *Indirect EOTP*, is for the sender to estimate the interference at the receiver by observing packet receptions and loss. That is, if the sender observes consecutive successful packet receptions, it assumes that the interference is low and reduces transmit power. If the sender observes a packet loss, it assumes that the interference is high and, therefore, increases transmit power immediately.

The other approach, which we call *Feedback EOTP*, is for the sender to obtain SINR information from the receiver using feedback. That is, the sender adjusts subsequent transmit power according to either the average or the minimum of a series consecutive SINR feedbacks.

Obviously, the former approach does not need feedback from the receiver and easier to implement, while the later needs feedback but reacts to interference variation faster. We implemented both versions of EOTP. Because they employ the same essential idea, their performance is similar in our subsequent simulations. For this reason, we only introduce Indirect EOTP in detail and, from now on, when mentioning EOTP, we mean Indirect EOTP.

5.2.3 Adjusting Transmit Power

In EOTP, the sender estimates the interference at the receiver by observing packet reception and loss. If the sender observes several consecutive successful packet receptions, it assumes that the interference is low and reduces P_{eotp}, where P_{eotp} is the transmit power used by EOTP. If the sender observes a packet loss, it assumes that the interference is high and, therefore, increases P_{eotp} immediately.

Before presenting the details of the power adjustment scheme, we consider the lower and upper bounds on P_{eotp}. It is clear that P_{min} is a lowerbound on P_{eotp}. Otherwise, the signal could not be correctly received even there was no interference. P_{otp} is an upperbound on P_{eotp}, because, when the sender-receiver distance is less than a threshold, P_{otp} guarantees reliable signal reception even when the receiver experiences the worst case interference. Both P_{min} and P_{otp} can be derived by exchanging control packets, such as RTS and CTS. Given the signal transmit power, P_t, and receive power, P_r, P_{min} can be calculated by the following equation

$$P_{min} = \frac{P_t}{P_r}\mathcal{T}_{rx} \ , \tag{5.1}$$

where $\mathcal{T}_{rx}$ is the receiving threshold. The OTP calculation, e.g. (4.8) in Chapter 4, can be rewritten as a function of P_t and P_r as well. It worth mentioning that P_{max} is also an upperbound on P_{eotp}. It is feasible to replace P_{otp} with P_{max} in subsequent derivations. But, in doing that, P_{eotp} would vary in a wider range.

Since P_{min} and P_{otp} define the lowbound and upperbound of P_{eotp}, we consider P_{eotp} to be a linear combination of P_{min} and P_{otp}, e.g.

$$P_{eotp} = k \times P_{otp} + (1 - k) \times P_{min} \ , \tag{5.2}$$

where k is a factor between 0 and 1 that reflects the interference level at the receiver. For example, if there is no interference, k is 0 and $P_{eotp} = P_{min}$; if there is significant

interference around the receiver, k is 1 and $P_{eotp} = P_{otp}$.

To guarantee reliable signal reception, k is varied based on a strategy of conservative decreases and aggressive increases. The sender sends m data packets at the current power level, P_{eotp}. If all the m data packets are successfully transmitted without retransmission, the sender reduces k by a small amount

$$k = max\{k - \delta, \quad \Delta\}, \tag{5.3}$$

where δ is the amount k is reduced each time, and Δ is slightly greater than 0 to guarantee that, at a minimum, P_{eotp} is slightly higher than P_{min}. When a retransmission occurs, k is increased immediately

$$k = min\{k + \delta, \quad 1\} \ . \tag{5.4}$$

Considering that the IEEE 802.11 allows at most 7 retransmissions, we choose a value for δ such that k would be likely to be 1 on the 7th retransmission. For example, setting δ to 0.1. Note that in both (5.3) and (5.4), "=" denotes assignment as in a computer program, not mathematical equality. By convention, the right-hand side of the "=" is called the R-Value and the left-hand side of the "=" is called the L-Value. R-values are based on the values of the terms before the assignment is completed, while L-values indicate what the term will be after the assignment. Thus, both (5.3) and (5.4) indicate that the new value of k is obtained by substituting the current value of k into the right-hand formula.

We claim that it is not desirable for the sender to increase P_{eotp} for an RTS retransmission or RTS/CTS exchange failure. This is because an RTS/CTS exchange failure occurs in two situations. One is that either the RTS or CTS is corrupted by interference. Another is that the receiver receives the RTS correctly but is not able to reply with a CTS because it is in the CSR of an ongoing transmission.

In the former situation, P_{eotp} should be increased, but, in the latter situation, P_{eotp} should not. Since there is no way to distinguish these two cases, we choose not to increase P_{eotp} when an RTS/CTS exchange fails.

5.2.4 The Transmit Power of Control Packets

In the previous discussion, we focus on data packet transmit power. The transmit power of control packets is important as well. We choose to send control packets at P_{eotp}, which is the same power level used by data packets, for several reasons.

First, if P_{eotp} guarantees data packets reception, it is likely to guarantee control packet reception, because the transmission duration of a control packet is much shorter than that of a data packet. Second, sending control packets at a higher power generates burst interference to ongoing transmissions, resulting in the senders in the vicinity increasing their transmit power. This effect may propagate to the whole network and degrade overall network performance. Third, sending control packets at a power level lower than P_{eotp} is unnecessary, because it would not reduce the interference envelop at the active receivers.

As in most MAC layer protocols, EOTP uses acknowledgements to realize reliable transmission. Therefore, the reliable reception of an ACK at the sender is equally important as that of a data packets at the receiver. This suggests that the receiver needs to track the interference level at the sender and adjusts its transmit power accordingly, just as the sender does. EOTP solves this issue indirectly such that the corruption of an ACK would result in data packet retransmission and thus increasing transmit power at both the sender and the receiver.

5.2.5 EOTP and Rate Adaptation

Similarly, the sender can adapt the transmission rate by observing packet losses and receptions, namely *Enhanced OTR* (EOTR). But, in doing this, the sender may

lose chances to reduce transmit power. As discussed in Chapter 4, the answer to the question of which scheme, EOTP or EOTR, brings about more significant improvement on network performance depends on specific traffic pattern and network topology and is left for future work.

5.3 Details of the EOTP Protocol

We first describe the packet format modifications and additional data structures in each node, then present the steps for EOTP, and finally discuss some feasibility issues.

5.3.1 Packet Format and Data Structure

We add a power level field, "PL", in the RTS and data packet header. Before sending an RTS or data packet, the sender sets "PL" to the current transmit power level, P_{etop}, and, therefore, notifies the receiver what power level it should use when replying with a CTS or an ACK.

Additionally, each sender maintains a power table with one entry per receiver containing five fields:

1. Receiver ID.

2. The minimal required transmit power, P_{min}, which is the lowerbound of P_{eotp}. Initially, this is set to P_{max}. To cope with node mobility, P_{min} is updated according to the transmit and receive power levels everytime a RTS/CTS exchange is completed.

3. The OTP transmit power, P_{otp}, which is the upperbound of P_{eotp}. Initially, this is set to P_{max}. As with P_{min}, P_{otp} is updated according to the transmit and receive power levels everytime a RTS/CTS exchange is completed so as to cope with node mobility.

4. A counter, CST, which counts the number of consecutive successful data packet transmissions. Initially, it is set to 0. CST is incremented when an ACK is received and it is reset to 0 when the factor k is changed.

5. The factor k. Initially, this is set to 1. k is reduced when CST reaches a threshold, CST_{th}, and it is increased when a data retransmission occurs.

In a practical system, there should be a timeout field indicating whether this information is outdated. We did not implement this functionality, because our subsequent simulations use traffic flows that continuously transmit packets and thus keep the entry fresh.

5.3.2 Details of the EOTP protocol

The steps of EOTP are described as follows:

1 Before sending an RTS, the sender looks up the power table entry of the intended receiver. If no entry is found, it adds one for the receiver and sets all the fields to their initial values. The sender calculates P_{eotp} using (5.2). Thereafter, it stores P_{eotp} into the "PL" field in the RTS header and sends the RTS at P_{eotp}.

2 The receiver receives the RTS, retrieves P_{eotp} from the header and replies with a CTS at P_{eotp}.

3 After receiving the CTS, the sender measures the receive power, updates P_{min} using (5.1) and updates P_{otp} using a variation of (4.8). Thereafter, the sender re-calculates P_{eotp}, stores P_{eotp} into the "PL" field in a data packet header and transmits the data packet at P_{eotp}.

4 The receiver receives the data packet, retrieves P_{eotp} and replies with an ACK at P_{eotp}.

5 Upon receiving the ACK, the sender increments CST. If CST reaches the threshold CST_{th}, k is reduced according to (5.3) and CST is reset to 0. Thereafter, the sender may go back to step 1 for the next round of data transmission.

6 If the sender does not receive an ACK from the intended receiver in a predefined time period, k is increased immediately according to (5.4) and CST is reset to 0.

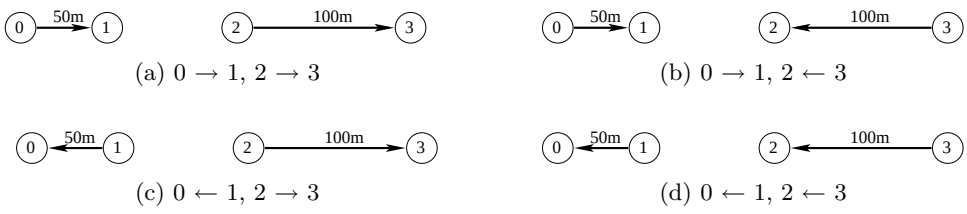
5.3.3 EOTP and Signal Power Estimation

In step 3, the sender updates P_{min} and P_{otp} using the transmit and receive power level of the control packets (either the RTS or the CTS). If positioning techniques (Section 2.10) are provided at each wireless node, we can obtain the sender-receiver distance, d, and step 3 can be simplified to calculating P_{min} and P_{otp} using the large-propagation model.

If a positioning technique is not available, the issue that needs to be addressed is for the receiver to separate the signal power, P_r, from interference, I, and white noise, N_0. As we discuss in Section 2.8, some sophisticated SINR estimators are able to fulfill this task. Therefore, if the receiver is capable of estimating SINR, step 3 is feasible.

Furthermore, if a SINR estimator is not provided at the receiver, we have to use the signal power approximations introduced in Section 2.9. In doing this, some errors will be introduced into P_{min} and P_{otp}. But it would not impact the essential idea of EOTP. For the worst case, we can let P_{eotp} vary between P_{max} and 0 and the performance of $EOTP$ may degrade. However, such extreme choices can be avoided, if we use the lower bound in (2.13) or (2.14) to overestimate P_{otp} and use the upper bound to underestimate P_{min}.

Thus, using the proposed techniques, step 3 is feasible.



(a) $0 \rightarrow 1$, $2 \rightarrow 3$ (b) $0 \rightarrow 1$, $2 \leftarrow 3$

(c) $0 \leftarrow 1$, $2 \rightarrow 3$ (d) $0 \leftarrow 1$, $2 \leftarrow 3$

Figure 5.2: Two traffic flows with four combinations of flow directions.

5.4 Simulation Results

We implemented EOTP by modifying the IEEE 802.11 DCF MAC protocol in NS2 [5, 6]. In our implementation, the CST threshold, CST_{th}, is set to 10 so that there is enough historical information for the sender to adjust subsequent transmit power. The two parameters in (5.3) and (5.4), e.g. Δ and δ, are both set to 0.1. The other parameters of our simulation setup are identical to those in Chapter 4.

Notice that (5.2), (5.3) and (5.4) are not the only functional forms of tuning P_{etop}. Likewise, there are other choices for CST_{th}, Δ and δ. We leave the exploration of the optimal functions and values for future work.

We used three network topologies: two traffic flows, chains, and grids. Since our goal is to evaluate the performance of MAC power control schemes, we eliminate the impact of routing by setting up static routes. Performance is evaluated in two ways: overall network throughput and average packet delivery latency. Packet delivery latency is the duration from the time a packet generated at a source to the time it is received by a destination. Because the packet delivery latency in two traffic flows and chain topologies is not very interesting, we present it only for the grid topologies. Our simulation demonstrates that EOTP outperforms both OTP and the IEEE 802.11 substantially.

5.4.1 Two Traffic Flows

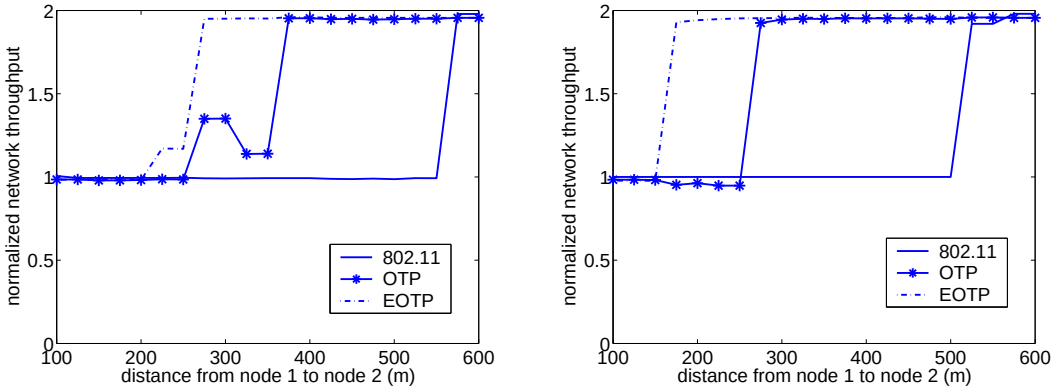
The two traffic flow topology is the simplest one for studying "inter-flow" spatial reuse. Fig. 5.2 shows four combinations of flow directions. The two traffic flows are Tr_{01} and Tr_{23} with the sender-receiver distance $50m$ and $100m$ respectively. Both senders inject CBR traffic at the full channel capacity. We varied the distance between node 1 and 2, d_{12}, from 100m to 600m and measured the network throughput with respect to different power control schemes.

Performance Comparison Between EOTP, OTP and 802.11

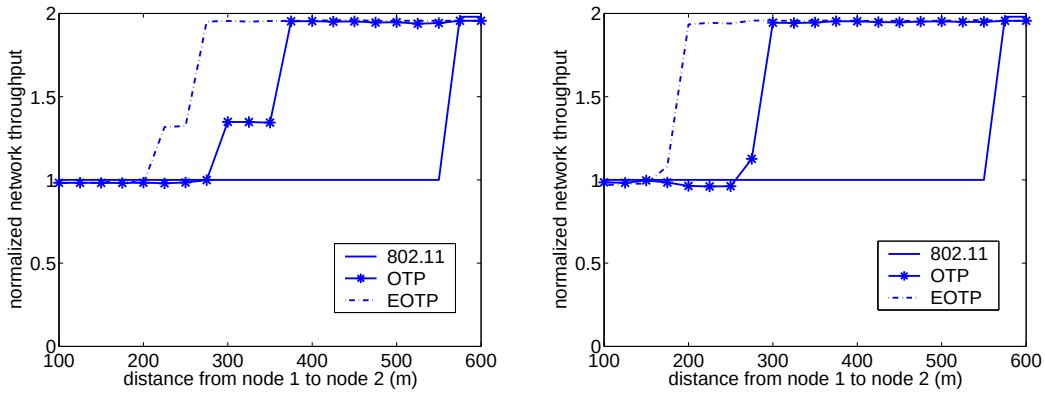
Figure 5.3 shows the simulation results for EOTP, OTP and 802.11. The x-axis is the distance from node 1 to node 2, and the y-axis is the network throughput normalized by the channel capacity. As we expected in all cases, the two traffic flows decouple from each other much earlier when employing EOTP than when employing either OTP or 802.11. This is because EOTP adjusts transmit power according to the actual interference level at the receiver, while OTP always assumes the worst interference at the receiver and transmits at a power level that guarantees successful packet reception in such situation. In this simulation, neither traffic flow generates much actual interference in the other flow. Thus, OTP is overly conservative. The explanation of the performance at each sample point is similar. We choose the case where $d_{12} = 300m$ in Figure 5.3(a) to explain the behavior of 802.11, OTP and EOTP in more detail.

- 802.11: Both node 0 and node 2 transmit at P_{max}, which is $0.2818w$ (24.5 dBm), and the radius of their CSRs, R_{cs}s, are $550m$. Each traffic flow is in the CSR of the other and so the two flows alternate transmissions. Therefore, the overall network throughput is the channel capacity.

- OTP: Node 0 and node 2 transmit at $0.0188w$ (12.7 dBm) and $0.0557w$ (17.5



(a) $0 \rightarrow 1$, $2 \rightarrow 3$ (b) $0 \rightarrow 1$, $2 \leftarrow 3$



(c) $0 \leftarrow 1$, $2 \rightarrow 3$ (d) $0 \leftarrow 1$, $2 \leftarrow 3$

Figure 5.3: Normalized network throughput for three MAC power control schemes: 802.11, OTP and EOTP.

dBm), and the R_{cs}s are $280m$ and $366m$ respectively. When node 0 is transmitting, node 2 can transmit simultaneously because node 2 is not in the CSR of node 0. However, when node 2 is transmitting, node 0 has to keep silent because it is in the CSR of node 2. Therefore, the overall network throughput is somewhat higher than the channel capacity.

- EOTP: When first beginning, both node 0 and node 2 set $k = 1$. As for OTP, their transmit power levels are $0.0188w$ (12.7 dBm) and $0.0557w$ (17.5 dBm), and the R_{cs}s are $280m$ and $366m$ respectively. During this stage, when node 0 is transmitting, node 2 can initiate simultaneous transmissions because it is out of the CSR of node 0. Node 2's transmission would not cause the worst possible interference at receiver 1 because its transmit power is much less than P_{max}. When node 2 is transmitting, node 0 has to keep silent because it is in the CSR of node 2. Therefore, receiver 3 detects no interference. Hence, both node 0 and 2 decrease their transmit power after 10 successful data packet transmissions. This process continues until both senders decrease their transmit power to the minimal level, which is $0.003181w$ (5 dBm) and $0.012742w$ (11.1 dBm) and the R_{cs}s are $180m$ and $255m$ respectively. At this point, neither node is in the CSR of the other. The two traffic flows are completely decoupled from each other and the overall network throughput is twice the channel capacity.

Evaluating the Two Enhancements of EOTP

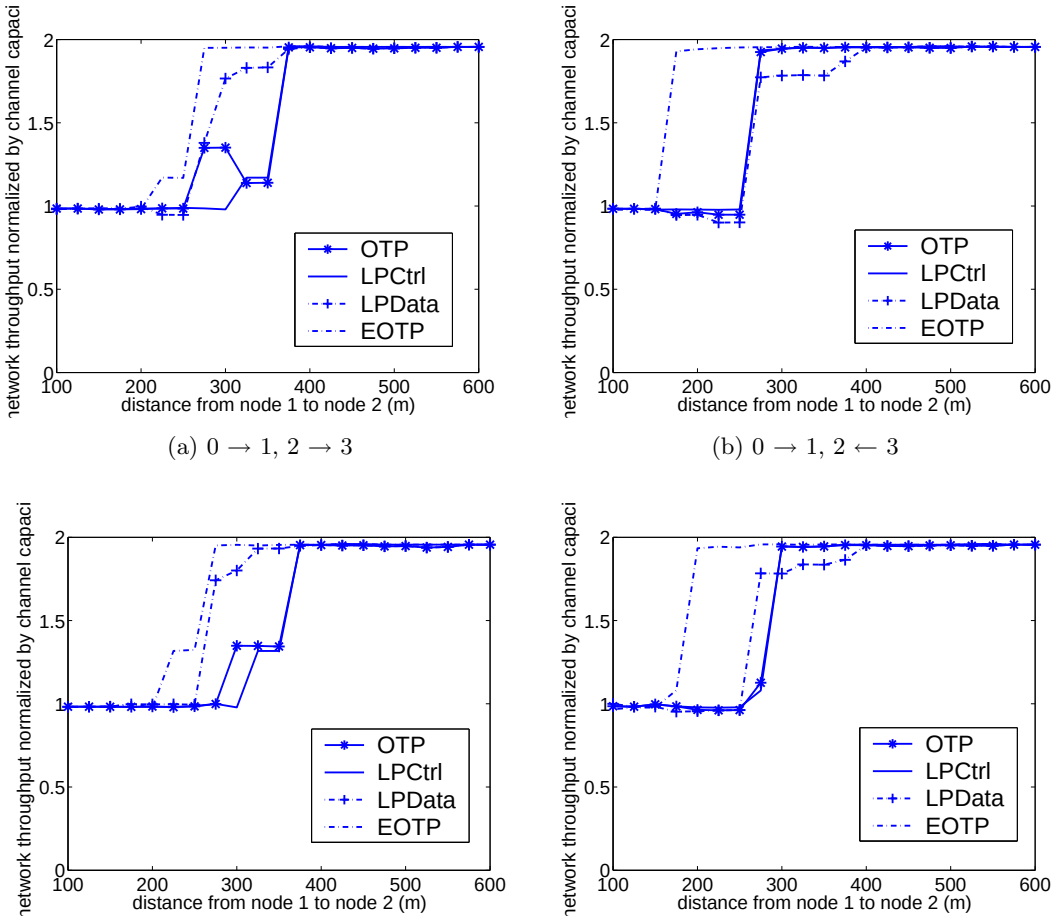
EOTP enhances OTP in two ways. One is reducing data packet transmit power, and the other is reducing control packet transmit power. We evaluate each enhancement independently by measuring the performance of two variations of OTP. In the first variation, data packets are transmitted at P_{eotp} while control packets are transmitted at P_{max}, denoted as LPData (*Low-Power Data packets*). In this

case, data packets are more sensitive to interference at the receiver. Sending control packets at P_{max} would increase the interference envelope at the receiver and prevent P_{eotp} from dropping to a low level. Thus, we would not expect LPData to improve the performance of OTP substantially. In the second variation, both data packets and control packets are transmitted at P_{otp}, denoted as LPCtrl (*Low-Power Control packets*). Because P_{otp} guarantees data packet receptions even when the receivers experience the worst possible interference, reducing transmit power of control packets has a trivial impact on data packet reception. Thus, we would not expect LPCtrl to improve the performance of OTP either.

Figure 5.4 illustrates the simulation results for EOTP, OTP and the two variations. The x-axis is the distance from node 1 to node 2, and the y-axis is the network throughput normalized by the channel capacity. As we expected in all cases, LPCtrl does not have a substantial impact on OTP. Although sending control packets at reduced power reduces the interference at the receivers, data packets are transmitted at P_{otp} so that successful packet reception is not sensitive to the actual interference level at the receiver. LPData also does not improve the performance of OTP in Figure 5.4(b) and (d), because sending control packets at P_{max} prevents the two traffic flows from decoupling from each other before $250m$ due to virtual carrier sensing. However, LPData improves OTP in Figure 5.4(a) and (c). This is because, when the flow distance is beyond $250m$, the two traffic flows can initiate more simultaneous transmissions due to the reduced data transmit power.

As in the previous simulation, we choose the typical case of $d = 300m$ in Figure 5.4(a) to explain the behavior of LPCtrl and LPData.

- LPCtrl: Node 0 transmits both control packets and data packets at $0.0188w$ (12.7 dBm) and thus the R_{cs} $280m$. Node 2 transmits both control packets and data packets at $0.0557w$ (17.5 dBm) and thus the R_{cs} $366m$. When node 0 is transmitting, node 2 can transmit simultaneously because node 2 is not



(a) $0 \rightarrow 1,\ 2 \rightarrow 3$

(b) $0 \rightarrow 1,\ 2 \leftarrow 3$

(c) $0 \leftarrow 1,\ 2 \rightarrow 3$

(d) $0 \leftarrow 1,\ 2 \leftarrow 3$

Figure 5.4: Evaluating the two enhancements of EOTP individually.

in the CSR of node 0. However, when node 2 is transmitting, node 0 has to keep silent because it is in the CSR of node 2. Assume that node 0 detects the channel to be idle and sends an RTS to node 1. Since the RTS is transmitted at $0.0188w$ (12.7 dBm), node 2 cannot sense it and attempts to reserve the channel, which prevents node 1 from replying with a CTS to node 0. When the traffic load at node 2 is high, node 1 has few chances to obtain the channel and thus the throughput for Tr_{01} is almost 0. This is an unfairness phenomena that occurs occasionally in IEEE 802.11 based networks. Therefore, the normalized network throughput is 1.

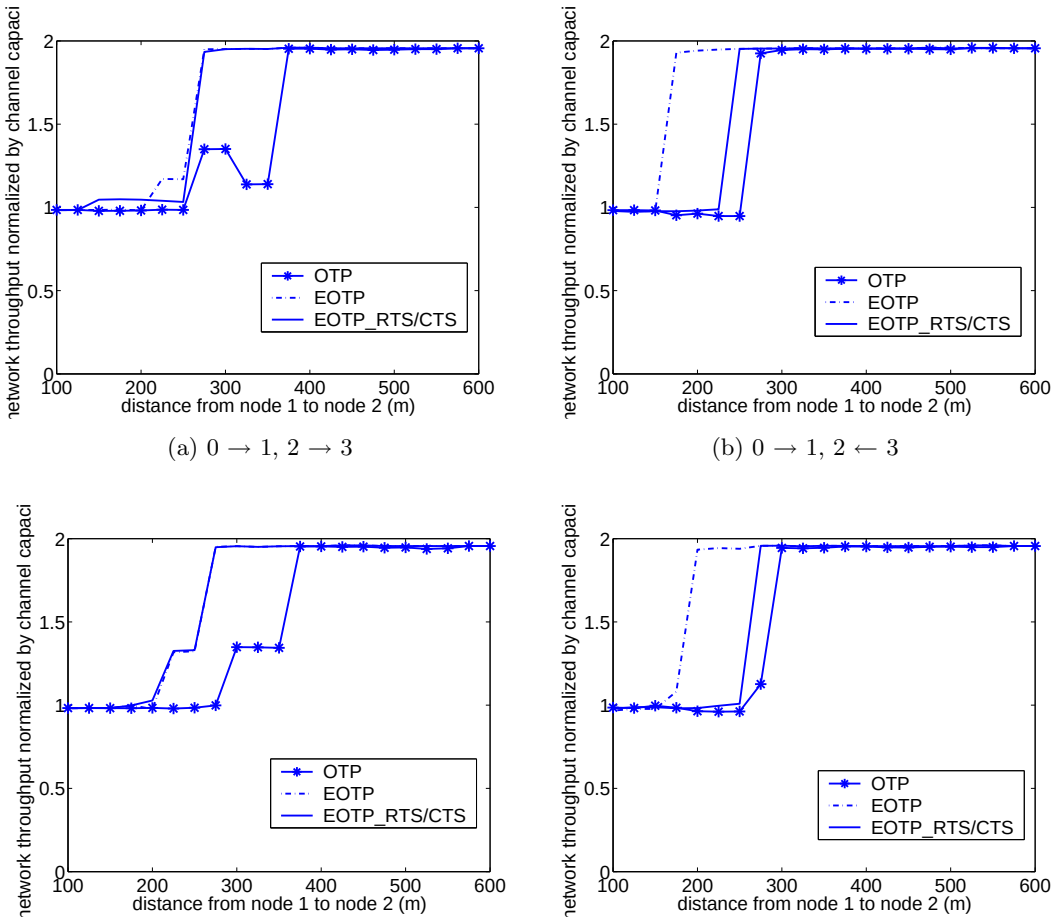
- LPData: The power adjustment process is the same as in EOTP. However, because control packets are transmitted at P_{max}, the interference envelope at both receivers is high. Node 0's data packet transmit power converges in the range of $0.0049w \sim 0.0066w$ (6.9 dBm $\sim$ 8.2 dBm) and node 2's converges in the range of $0.0127w \sim 0.0272w$ (11 dBm $\sim$ 14.3 dBm). Because of the variation of the transmit power, traffic flow Tr_{01} and Tr_{23} couple with each other occasionally. The overall network throughput is lower than two times channel capacity.

Thus, we conclude that, although each enhancement has limited impact on OTP, their synergistic effect is significant.

Power Adjustment on RTS/CTS Exchange Failure

To test our design rationale about whether or not to increase transmit power for RTS/CTS exchange failure, we evaluate the performance of a variation of EOTP where the sender increases transmit power not only for data packet loss but also when the RTS/CTS exchange fails. This scheme is denoted as EOTP_RTS/CTS.

Figure 5.5 illustrates the simulation results using the same format as previous Figures. Figure 5.5(b) and (d) show that EOTP_RTS/CTS underperforms



(a) $0 \rightarrow 1$, $2 \rightarrow 3$

(b) $0 \rightarrow 1$, $2 \leftarrow 3$

(c) $0 \leftarrow 1$, $2 \rightarrow 3$

(d) $0 \leftarrow 1$, $2 \leftarrow 3$

Figure 5.5: Performance comparison between EOTP and EOTP_RTS/CTS.

EOTP. This is because, in most cases, RTS/CTS exchange fails because the receiver is reserved by other ongoing transmissions and is not able to reply with a CTS. Increasing transmit power cannot alleviate this problem but, instead, causes the two traffic flows to couple together. We use the typical case of $d_{12} = 200m$ in Figure 5.5(b) to explain the behavior of EOTP and EOTP_RTS/CTS.

- EOTP_RTS/CTS: When first beginning, both node 0 and node 3 set $k = 1$ and their transmit powers are $0.0188w$ (12.7 dBm) and $0.0557w$ (17.5 dBm) respectively. Because neither flow generates much interference in the other, both senders gradually decrease their power levels. When node 3 reduces its power level to $0.0416w$ (16.2 dBm) and the R_{cs} $340m$, the CSR covers node 1 but excludes node 0. The RTS/CTS exchange between node 0 and 1 fails frequently because, most of the time, node 1 is reserved by the ongoing transmissions of Tr_{23}. Thus, node 0 increases the transmit power all the way back to the upperbound $0.0188w$ (12.7 dBm), which, in turn, frequently reserves receiver 2, resulting in node 3 continuously increasing its power to $0.0557w$ (17.5 dBm). This process repeats until the end of the simulation. As a consequence, node 0 transmits at $0.0188w$ (12.7 dBm) most of the time and node 3 varies from $0.0368w$ (15.7 dBm) to $0.0557w$ (17.5 dBm). The two flows couple with each other and the network throughput is the channel capacity.

- EOTP: Both senders do not increase transmit power when the RTS/CTS exchange fails. Therefore, in the situation above, both node 0 and node 3 keep reducing their power levels as long as no data packet retransmission occurs. Eventually, both senders decrease their transmit power to the minimum level, resulting in the two flows decoupling from each other. Thus, the network throughput achieves twice the channel capacity.

In the scenario illustrated by Figure 5.2(a), receiver 1 may be silenced by flow

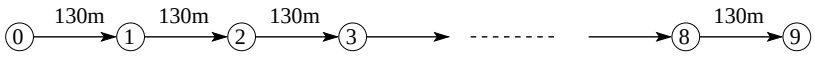


Figure 5.6: The chain topology

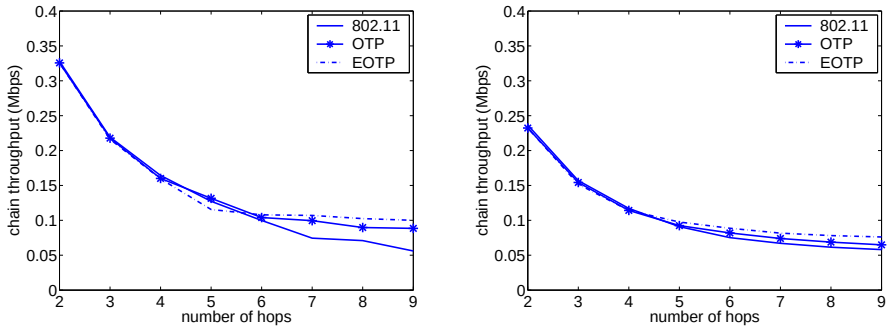
Tr_{23} and unable to reply with a CTS to sender 0, resulting in EOTP_RTS/CTS increasing the power level at node 0. But, this would never occur at receiver 3 because it is further away from flow Tr_{01} than sender 2. Eventually, node 2 would decrement its power to the minimum level, leaving node 1 outside of its carrier sense range. At this moment, node 0 could gradually decrease its power level. As a result, EOTP_RTS/CTS demonstrates similar performance to EOTP, as shown in Figure 5.5(a). In the scenario illustrated by Figure 5.2(c), both receivers are further way from the other flow than the senders. Thus, higher interference would be the only reason that causes an RTS/CTS exchange failure. In this scenario, EOTP_RTS/CTS acts as EOTP and their performance are the same, as shown in Figure 5.5(c).

In a conclusion, EOTP_RTS/CTS does not outperform EOTP in any case and it does underperform EOTP in some. Therefore, it is appropriate to not increase transmit power for the RTS/CTS exchange failure in EOTP.

5.4.2 Chain Topologies

Our second study focuses on chain topologies. The chain topology is the typical one for studying "intra-flow" spatial reuse. As shown in Fig. 5.6, 10 nodes reside on a straight line and the immediate neighbors are separated by 130m so that a node can only reach its immediate neighbors even if it transmits at P_{max}. In this way, EOTP, OTP and 802.11 would use an identical route to delivery data packets.

The source, node 0, attempts to inject packets at the full channel capacity. We vary the destination from node 2 to node 9, and measure the chain throughput. We exclude node 1 as the destination because, in this case, there is no interference



(a) cbr traffic flow

(b) tcp traffic flow

Figure 5.7: Throughput for the chain topology.

or channel contention. Moreover, the throughput of the single hop flow, which is around 0.7Mbps, is much higher than the others. Thus, an extra benefit of excluding node 1 is that we can use a small scale on the y-axis and distinguish the performance of the schemes.

The simulation results are shown in Figure 5.7, where the x-axis is the number of hops from node 0 to the destination and the y-axis is the throughput in Mbps. Figure 5.7(a) and (b) demonstrate the simulation results for CBR and TCP traffic respectively. EOTP slightly outperforms OTP and 802.11 in both cases. The improvement is trivial because all three protocols deliver data packets along the paths consisting of the least number of hops and, therefore, each hop distance can not be less than $d_{max}/2$, which limits the performance of MAC layer power control schemes. An addition explanation is that the chain topology is a 1D topology and Chapter 3 suggests that power control in 1D space may not be able to improve network throughput.

5.4.3 Grid Topologies

The previous simulations were performed using one dimensional topologies. To evaluate the performance of EOTP in two dimensional space, we use a 700m by

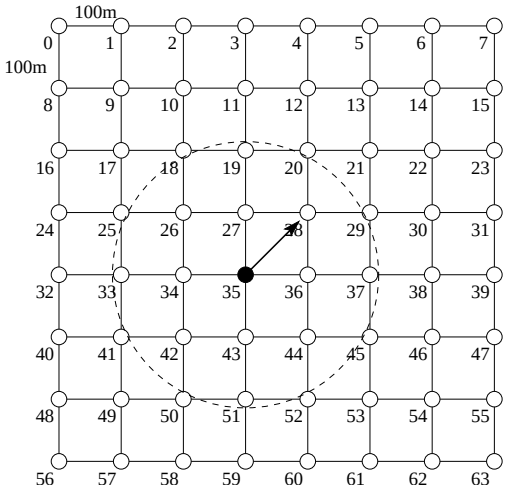


Figure 5.8: A 700m by 700m grid topology with neighboring nodes 100m apart.

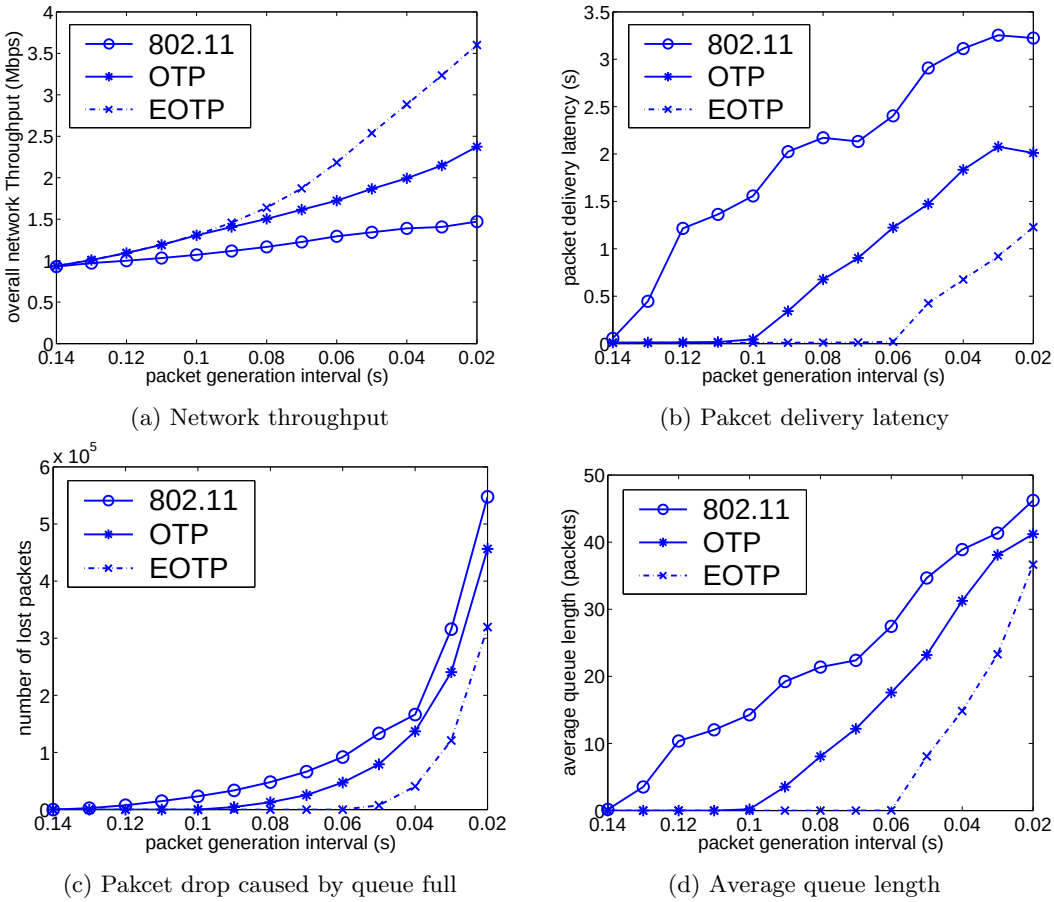
700m grid topology with neighboring nodes 100m apart, as shown in Figure 5.8. Each node is labeled on its left side.

To eliminate the impact of routing and focus on the MAC protocol performance, a node sends CBR traffic to a destination that is just a single hop away. In the subsequent study of routing strategy, we perform similar simulations except that there is no restriction on source-destination distance.

We test two sets of traffic flows. In the first set, the traffic flows are regularly distributed in the network, while, in the second set, the traffic flows are randomly distributed. To fully utilize the network spatial resource and, therefore, evaluate the power control schemes fairly, every node participates in communications. Specifically, in the first flow set, a node is either a sender or a receiver, and, in the second flow set, every node is a sender.

Regularly Distributed Traffic Flows

In the first traffic flow set, each even-numbered node sends CBR traffic to its immediate neighbor on the right so that the distance from a sender to a receiver is 100m



(a) Network throughput

(b) Pakcet delivery latency

(c) Pakcet drop caused by queue full

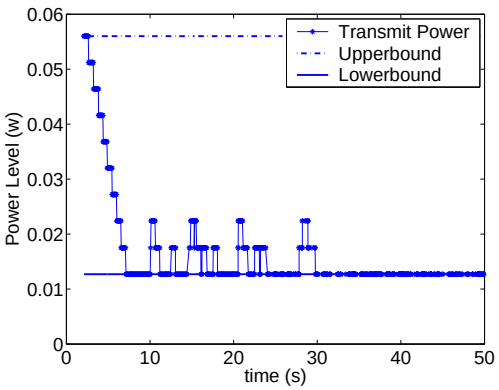
(d) Average queue length

Figure 5.9: Performance of EOTP, OTP and 802.11 in the grid topology with regularly distributed traffic flows.

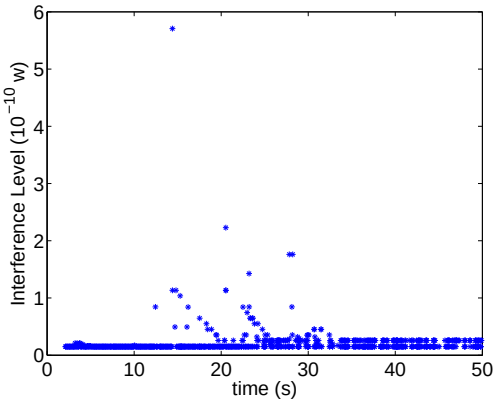
and the traffic load is evenly distributed in the network. Each traffic flow starts at a random time within the first 30 seconds and continues until the end of the simulation which lasts 500 seconds. To eliminate the impact of the initial start-up process, we measure the network throughput and packet delivery latency after the transmit power of all the source nodes converges. In this simulation, the time to achieve convergence is around 35 seconds.

Figure 5.9 shows the network throughput and packet delivery latency with varying traffic load. The x-axis is the packet generation interval at each source node. The smaller the packet interval, the heavier the load. It is clear that there is no need to reduce transmit power when network traffic load is light. Thus, we intentionally set the traffic load to be high enough to distinguish the performance of the MAC power control schemes. We observe that EOTP outperforms OTP and 802.11 substantially for both metrics.

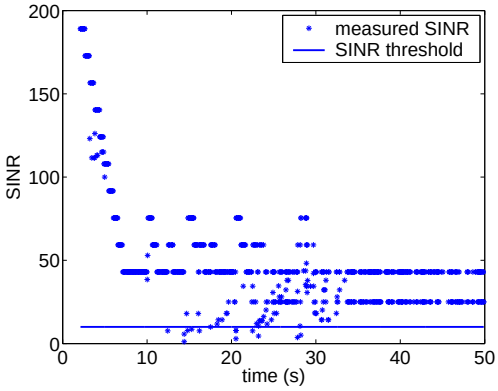
To illustrate the EOTP power adjustment process, we repeat the simulation in which the packet generation interval is 0.05 seconds, and observe the EOTP power adjustment process at node 26 as well as the interference variation at the corresponding receiver 27. Each time node 26 sends a data packet to node 27, we record the transmit power at node 26, the highest interference and the lowest SINR at node 27. Figure 5.10(a) illustrates the transmit power variations at node 26 during the first 50 seconds. The dotted line is the transmit power upperbound, $0.0557w$ (17.5 dBm), and the solid line is the transmit power lowerbound, $0.0127w$ (11 dBm). Figure 5.10(b) and (c) show the interference and SINR variations at node 27 for the first 50 seconds. The solid line in Figure 5.9(c) is the SINR threshold for capture effect, which is 10 (10 dB) in our simulation setup. As we can see, at the beginning of the simulation, the transmit power drops sharply from $0.557w$ (17.5 dBm) to $0.0127w$ (11 dBm), because few flows have started at this time. As other traffic flows start, burst interference causes node 26 increases its transmit



(a) Transmit power at node 26

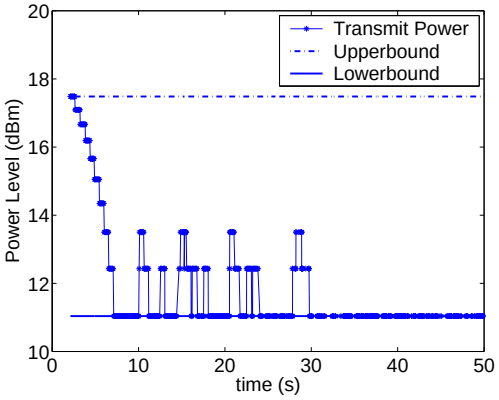


(b) Interference at node 27

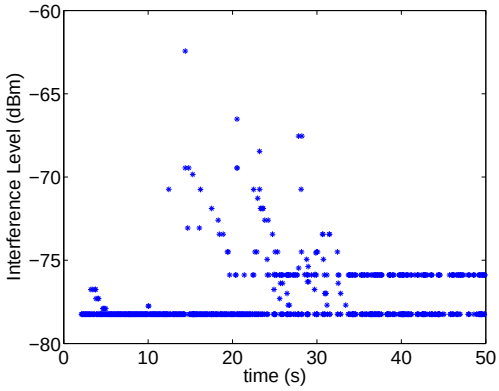


(c) SINR at node 27

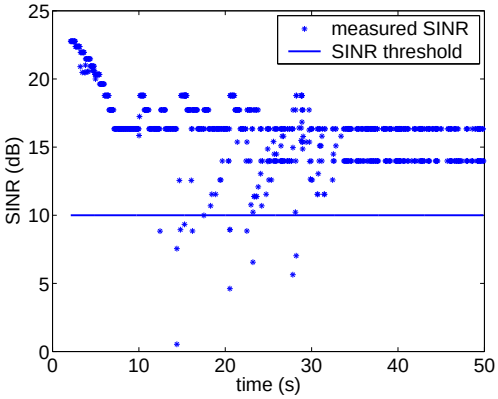
Figure 5.10: EOTP power adjustment process at node 26.



(a) Transmit power at node 26



(b) Interference at node 27



(c) SINR at node 27

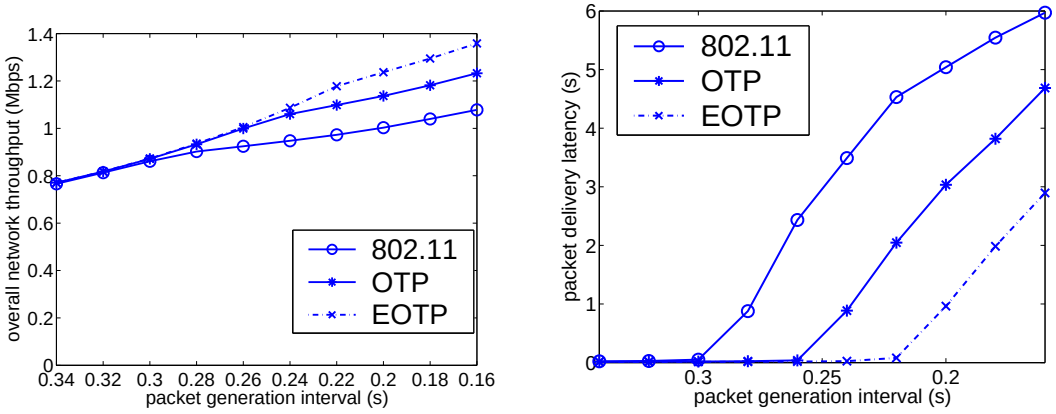
Figure 5.11: EOTP power adjustment process at node 26.

power occasionally. For example, at 12 seconds, burst interference corrupts a data packet transmission, resulting in the transmit power being increased at node 26. This process continues until about 30 seconds, when all traffic flows have started and most senders have converged their transmit power. The interference at node 27 is stable and the transmit power at node 26 converges to the minimum level.

Randomly Distributed Traffic Flows

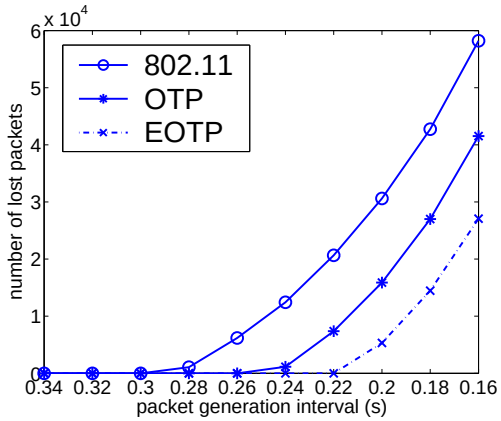
In the second traffic flow set, each node randomly chooses a destination that is less than $200m$ away and sends CBR traffic to it until the end of the simulation. During the simulation, these source-destination relationships do not change. Each traffic flow starts at a random time within the first 50 seconds. To eliminate the impact of the initial start-up process, we measure the network throughput and packet delivery latency after the transmit power of all the source nodes converge. Unlike the previous simulation, the duration of the convergence is significantly impacted by the traffic load. This is because there are a greater number of flows and the packet rate is lower in the second flow set than in the first one. It takes longer time for a source node to transmit enough number of data packets before determining its subsequent transmit power. The higher the traffic load the longer time it takes. Thus, the time to achieve convergence varies from 75 seconds to 200 seconds as packet interval decreases from 0.2 seconds to 0.06 seconds.

Figure 5.12 shows the simulation results from low to medium traffic load. Figure 5.12(a) is the overall network throughput in Mbps. The x-axis is the packet generation interval at each source node. The smaller the interval the higher the traffic load. We observe that when traffic load is low, the performance of EOTP, OTP and 802.11 is identical. This is because when traffic load is low, the spatial resource is sufficient for all transmissions and, therefore, power control is not helpful. However, as traffic load increases, EOTP outperforms both OTP and 802.11.

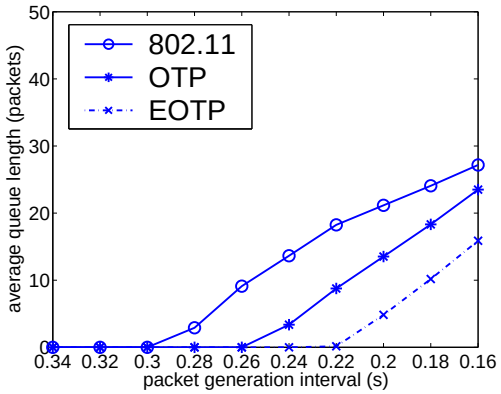


(a) Network throughput

(b) Packet delivery latency

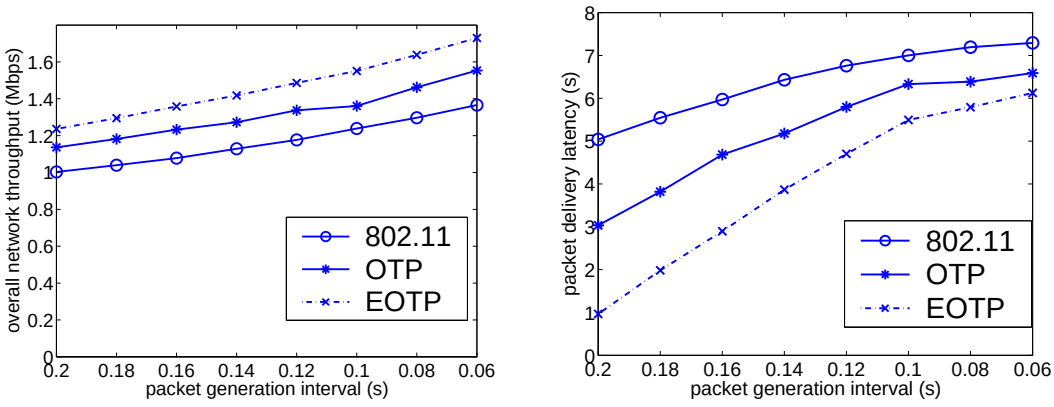


(c) Pakcet drop caused by queue full



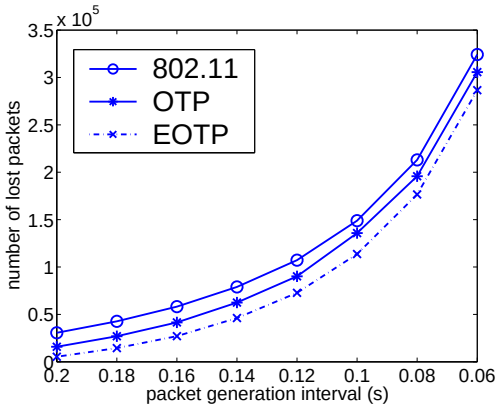
(d) Average queue length

Figure 5.12: Performance of EOTP, OTP and 802.11 in the grid topology with randomly distributed low traffic load.

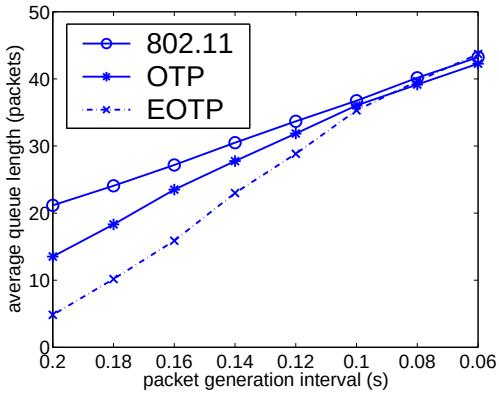


(a) Network throughput

(b) Packet delivery latency



(c) Packet drops caused by full queues



(d) Average queue length

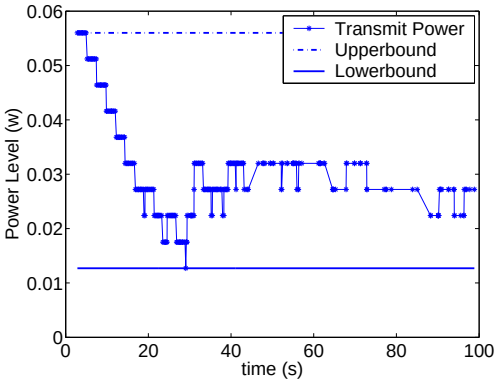
Figure 5.13: Performance of EOTP, OTP and 802.11 in the grid topology with randomly distributed high traffic load.

Figure 5.12(b) is the average packet delivery latency in seconds. When the traffic load is low, there is no channel contention and the three schemes have similar performance. When traffic load increases, channel contention increases. Power control reduces channel contention by reducing the space utilization of each individual transmission and, thus, significantly reduces queueing latency. This is confirmed by Figure 5.12(d), which shows the average queue length at each source node. It is obvious that, when the traffic load is high, the queue length of EOTP is much shorter than OTP and 802.11. In the cases that the EOTP has similar network throughput, for example when the packet generation interval is 0.28 seconds, its packet delivery latency has shown substantial improvement over 802.11 and OTP.

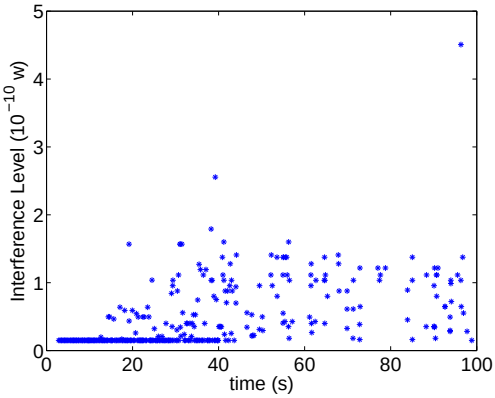
Figure 5.12(c) shows the number of lost packets caused by full queues. We observe that EOTP has lower packet loss than OTP and 802.11 when the traffic load is high. Notice that, due to the border effect, the nodes in the middle of the network are subjected to more severe channel contention than the nodes on the border. Thus, although the average queue length is much less than 50, which is the predefined queue size, packet loss is substantial.

Figure 5.13 shows the performance of EOTP, OTP and 802.11 when the traffic load is high. Figure 5.13(a) shows the network throughput in Mbps. The x-axis is the packet generation interval at each source node. The smaller the interval, the higher the traffic load. We observe that EOTP outperforms both OTP and 802.11.

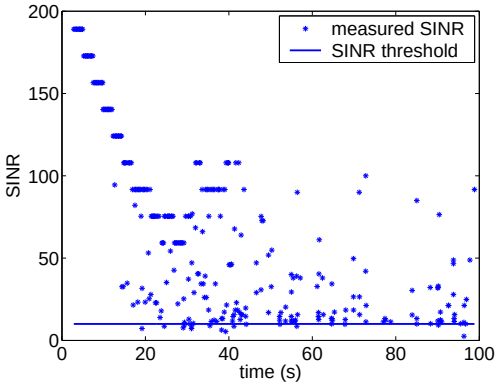
Figure 5.13(b) shows average packet delivery latency in seconds. EOTP has lower packet delivery latency than OTP and 802.11. We notice that as the traffic load increases, the performance gap decreases. This is because, when traffic load is sufficiently high, all three schemes have saturated queues and the packet delivery latency is high. This can be observed from Figure 5.13(c) and (d). When the packet generation interval is 0.06 seconds, all three schemes have similar average queue



(a) Transmit power at node 26



(b) Interference at node 27

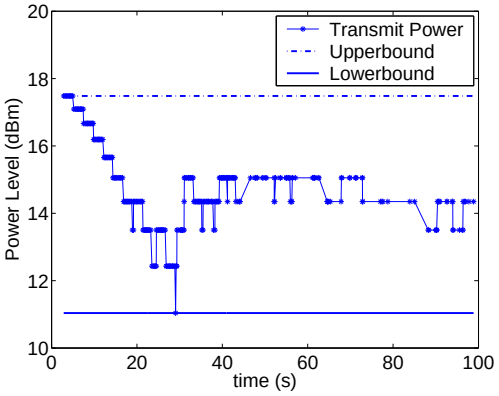


(c) SINR at node 27

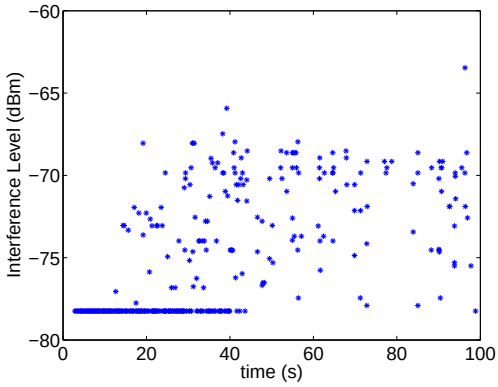
Figure 5.14: EOTP power adjustment process at node 26.

length but EOTP has the least number of dropped packets.

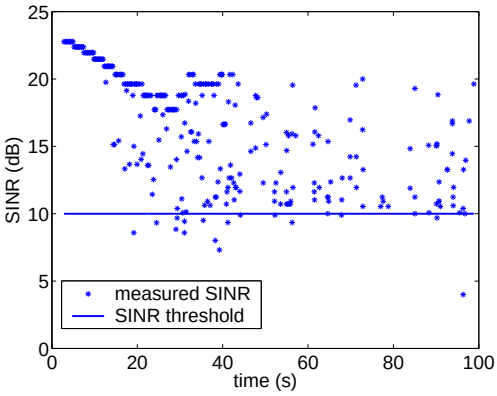
We also notice that the performance gap of the three schemes is less significant than that in Figure 5.9. This is caused by two reasons. First, the sender-receiver distances in the second traffic set, which are from $100m$ to $200m$, are longer than those in the first set, which are all $100m$. Therefore, in the second flow set, the senders transmit at higher power levels and generate more interference thus lowering the overall performance. Second, the traffic load in the second set is not evenly distributed, which may cause hot spots and performance degradation.



(a) Transmit power at node 26



(b) Interference at node 27



(c) SINR at node 27

Figure 5.15: EOTP power adjustment process at node 26.

Furthermore, we repeat the simulation in which the packet generation interval is 0.2 seconds, except that we manually force node 26 to send packets to node 27 so as to compare its power adjustment process to the previous simulation. Figure 5.14(a) illustrates the transmit power variations at node 26 for the first 100 seconds. The dotted line is the transmit power upperbound, $0.0557w$ (17.5 dBm), and the solid line is the transmit power lowerbound, $0.0127w$ (11 dBm). Generally speaking, the transmit power at node 26 converges to a range of $0.032w$ (15.1 dBm) to $0.022w$ (13.4 dBm). Figure 5.14(b) and (c) show the interference and SINR variations at node 27 for the first 100 seconds. The solid line in Figure 5.9(c) is the SINR threshold for capture effect, which is 10 in our simulation setup. As we can see, even after 50 seconds when all traffic flows have started and most source nodes have tuned their transmit power to appropriate levels, the interference at node 27 is much higher than that in the previous simulation. This is because some nodes around 27 have to transmit at a high power level to reach their destinations, which causes high interference at node 27. For example, at time 96 seconds, burst interference, which caused by a concurrent transmission from node 29 to node 13, corrupts the data packet at node 27, resulting node 26 to increase its transmit power immediately. The traffic flow from node 29 to node 13 starts at 30 second. It did not corrupt packet reception at node 27 earlier because both node 26 and node 29 have to compete the channel with other senders and they did not get the chance to transmit concurrently until 96 seconds. This observation confirms our motivation for enhancing OTP. That is, although the worst potential interferers exist, they may not be able to frequently corrupt packets at the receiver due to the low probability of obtaining the channel simultaneously. Thus, the sender can fruitfully risk transmitting packets at power levels lower than P_{otp}.

Less Aggressive Backoff

EOTP adjusts transmit power based on the strategy of "conservative decreases and aggressive increases". It specifies that a transmitter increases transmit power immediately after one packet loss. A consideration is that this might be overly aggressive. Therefore, we performed the same simulations as above to observe the performance if transmit power is increased only after n consecutive packet losses, where n is set to be 1, 2 and 4. Figure 5.16 shows the simulation results.

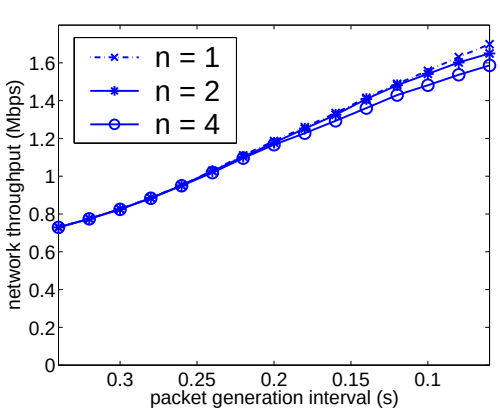
Figure 5.16(a) shows the network throughput in Mbps and the x-axis is packet generation interval in seconds. The smaller the generation interval, the higher the traffic load. We observe that, when traffic load is low, the performance is identical for $n = 1$, $n = 2$ and $n = 4$. However, when traffic load increases, the EOTP with $n = 1$ slightly outperforms the other two schemes.

Figure 5.16(b) shows the average packet delivery latency in seconds and the x-axis is packet generation interval in seconds. Again, we observe that, when traffic load is low, the performance is identical for $n = 1$, $n = 2$ and $n = 4$. However, when traffic load increases, the EOTP with $n = 1$ has a lower latency than the other two schemes.

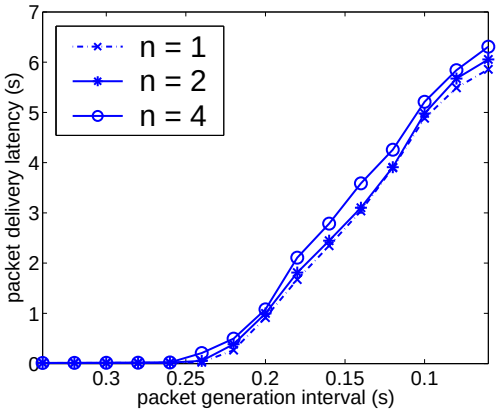
In this simulation, the performance of EOTP is not very sensitive to the value of n. We may obtain some interesting observations if using some other network topologies or traffic patterns. We leave it for future work.

5.4.4 Network Scale

As pointed out in Chapter 3, network scale impacts the performance of MAC power control schemes because of the border effect. Here, network scale is defined as the ratio between network size and sender-receiver (S-R) distance of individual transmissions. When network scale is small, both EOTP and OTP cannot improve network performance. We use simulations to demonstrate.



(c) Network throughput

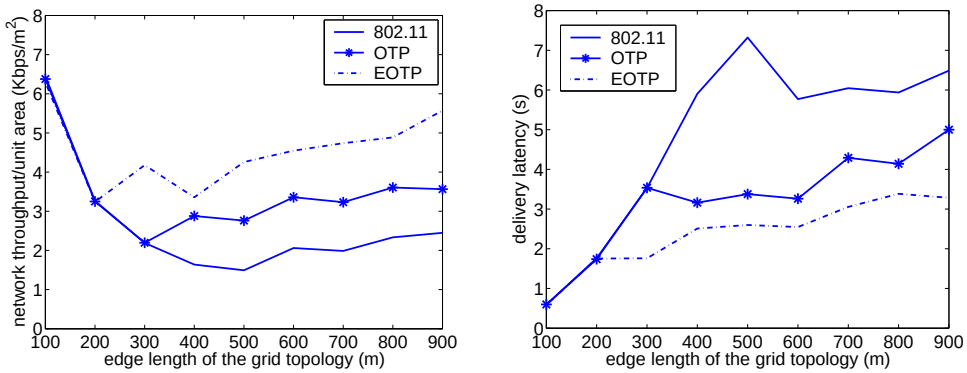


(d) Packet delivery latency

Figure 5.16: Increasing transmit power on different number of packet loss.

Network Size

First, we investigate the impact of network size on the performance of MAC power control schemes by keeping neighboring nodes 100m apart and varying the size of a grid topology from 100m by 100m, 200m by 200m to 900m by 900m. Thus, we vary network scale by fixing the S-R distance and changing the network size. To measure the achievable network throughput, each node sends CBR traffic to its immediate neighbor on the right at the full channel capacity. Figure 5.17(a) shows the network throughput for 802.11, OTP and EOTP. The x-axis is the edge length of the grid topology. The y-axis is the network throughput per unit area in $Kbps/m^2$, which is the overall network throughput divided by the network area. We observe that, when the network size is small, neither EOTP nor OTP outperforms 802.11. This is because, in the small-sized networks, the spatial resource conserved by the power control schemes is not utilized by other transmissions. For example, in the 200m by 200m grid network, the S-R distances are all 100m. When employing 802.11, the R_{cs} of an individual transmission is $550m$. Although EOTP reduces the R_{cs} to $200m$, the conserved space mostly falls outside of the network and cannot be



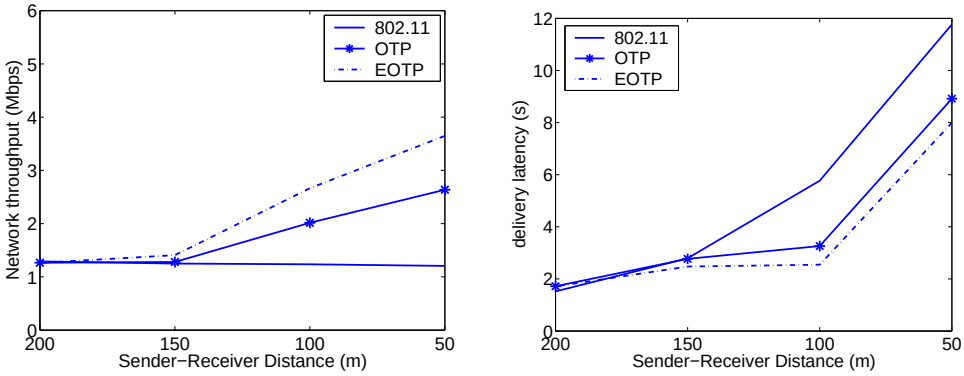
(a) Network throughput (b) Packet delivery latency

Figure 5.17: Network size impacts the performance of MAC power control schemes.

utilized. Figure 5.17(b) shows the average packet delivery latency. The x-axis is the edge length of the grid topology and the y-axis is the delivery latency in seconds. Also, we observe that, when the network size is small, power control schemes do not reduce packet delivery latency. This is because, the minimum sender-receiver distances are 100m and the transmit power cannot be reduced to a sufficient low level to decouple the traffic flows in the small-sized networks.

Sender-Receiver Distance

Second, we investigate the impact of the S-R distance on the performance of the MAC power control schemes by varying neighboring nodes distance from 50m, 100m, 150m to 200m in a 600m by 600m grid topology. Thus, we vary network scale by fixing the network size and changing the S-R distances. The shorter the S-R distance, the larger the network scale. To measure the achievable network throughput, each node sends CBR traffic to its immediate neighbor on the right at full channel capacity. Figure 5.18(a) illustrates the network throughput in Mbps. The x-axis is the S-R distance. When the S-R distances are long, neither EOTP nor OTP outperforms 802.11 because, in this situation, the transmit power could not be reduced



(c) Network throughput (d) Packet delivery latency

Figure 5.18: Sender-Receiver distances impact the performance of MAC power control schemes.

substantially due to the minimal required power level. For example, when the S-R distances are 200m, neither EOTP nor OTP could reduce the transmit power to a sufficiently low level to decouple the traffic flows. Figure 5.18(b) shows the average packet delivery latency in seconds. The x-axis is the S-R distance. For the same reason, when the S-R distances are long, both EOTP and OTP could not improve packet delivery latency. This simulation also suggests that MAC layer power control schemes favor short S-R distances.

Chapter 6

Mini-Hop Routing

Chapter 3, 4 and 5 suggest that short hop-distance benefits the performance of MAC power control schemes because individual transmissions can reserve less space and generate less interference. Mini-hop routing is a strategy that discovers paths consisting of short hops. Unlike shortest-path routing, which attempts to discover paths consisting of the least number of hops, mini-hop routing discovers paths that contain a greater number of shorter distance hops. Even though there is a tradeoff with the number of hops, we will show that the combination reduces the spatial usage of delivering each individual packet and therefore improves overall network throughput and, perhaps more surprisingly, packet delivery latency.

We use the spatial usage metric Ω and an example to explain why short hops may benefit network performance. We have shown in Chapter 3 that, in 2D space,

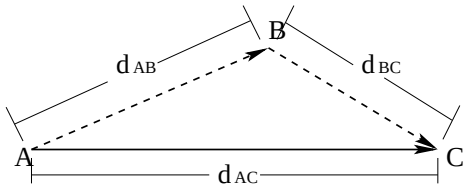


Figure 6.1: Shortest-path routing vs. mini-hop routing.

if routing control messages and data packets are sent at power level P_k, and thus transmit range R_k, n_h is linearly proportional to $\frac{1}{R_k}$, while a_i is proportional to R_k^2. Therefore, overall Ω is proportional to R_k. The smaller the transmit range, the lower the spatial usage to deliver each individual packet. Also remember we showed that, in 1D space, short hop do not improve spatial usage.

Figure 6.1 illustrates an example, where node A is the source and node C is the destination. A can reach C through the shortest path which contains a single hop "A→C" with distance d_{AC}. Also, it can go through two mini hops "A→B→C" with distances d_{AB} and d_{BC} respectively. If we employ a MAC layer power control scheme such that each hop transmission uses the minimal required power level, the spatial usage of forwarding a packet along "A→C" is $\Omega_1 = t\pi\gamma^2(d_{AC})^2$. Here, t is the packet transmission duration, and we ignore the number of retransmissions. Similarly, the spatial usage of forwarding a packet along "A→B→C" is $\Omega_2 = t\pi\gamma^2(d_{AB}^2 + d_{BC}^2)$. As long as $\angle ABC$ is greater than $\frac{\pi}{2}$, Ω_2 is less than Ω_1, which suggests that delivering packets along the mini-hop path, "A→B→C", consumes less spatial resource than delivering packets along the shortest path, "A→C".

In this Chapter, we propose a mini-hop routing protocol and demonstrate that this protocol, combined with an appropriate MAC power control scheme, improves network performance.

6.1 Related Work

Shortest-path routing discovers routes consisting of the least number of hops and, thus, each hop tends to span a long distance. To achieve efficient spatial utilization, mini-hop routing may be desirable so that packets can be delivered along routes consisting of a greater number of short hops.

Some mini-hop routing protocols have been proposed [112, 113, 114, 115]. Most of these protocols discover mini-hop paths by transmitting routing control

packets at reduced power levels, which is equivalent to setting an upperbound on hop-distance. In these schemes, the key issue is to identify the critical power, P_c, at which routing control packets should be transmitted. P_c should be low enough to exclude long distance hops and high enough to guarantee connectivity between sources and destinations. The proposed protocols identify P_c at the cost of substantial overhead.

In [113] and [114], routing control packets are transmitted at n different power levels during each individual route discovery process. Among the discovered paths, a source node chooses the one corresponding to the lowest power level. In the worst case, the routing overhead is n times higher than that of a normal route discovery.

In [112] and [115], each node periodically broadcasts beacons or hello messages at P_{max} on a control channel. A node maintains a connectivity set by overhearing these signals of its neighbors. The connectivity set is the smallest subset of its possible neighbors that guarantees connectivity of the node to the network. The node adjusts its transmit power to the level that can just reach the furthest node in the connectivity set. Thereafter, [112] employs table-driving routing such that each link is assigned a cost corresponding to the reduced power level and the routing protocol calculates the minimal cost path, while [115] employs on-demand routing such that each node simply transmits routing control packets at the reduced power level. Because of the beacon and hello messages, both protocols introduce considerable overhead.

6.2 Design Rationale

We propose a Mini-Hop Routing (MHR) protocol to discover paths consisting of short distance hops without introducing substantial routing overhead. The protocol design is based on the following considerations.

6.2.1 Assumptions

We base our work on several assumptions. First, we assume a single radio channel for communications instead of multiple channels. In doing this, we maintain hardware simplicity and avoid significant modifications to the physical layer of the IEEE 802.11 [1] which is a widely used WLAN MAC protocol.

Second, we assume that, given transmit and receive power levels, the distance between a sender and a receiver can be estimated according to the signal propagation model [8], as discussed in Chapter 2. Although this assumption is challenged by the existence of fast fading, we may solve it through more sophisticated physical layer solutions. For example, the sender and receiver can eavesdrop each other's transmissions and average their distance over a long interval. Moreover, the distance can be obtained through other techniques such as GPS or triangulation.

Third, we base MHR on Dynamic Source Routing (DSR) [35]. This is because DSR is a typical on-demand routing protocol and on-demand routing is considered to be more appropriate for multihop wireless networks [31]. Incorporating mini-hop routing into other on-demand routing protocols is essentially the same as incorporating it into DSR.

6.2.2 Two Approaches to Discover Mini-Hop Routes

As introduced in Chapter 2, typical on-demand routing protocols employ a blind flooding technique to discover routes and the blind flooding technique tends to discover long hops. There are two obvious methods to discover mini-hop routes during the blind flooding route discovery process: the power-control method and the link-cost method.

The power-control method employs power control in the process of route discovery. That is, each node transmits routing control packets at a reduced power level, P_c. P_c has to be low enough to exclude long distance hops and high enough

to maintain connectivity between sources and destinations. The advantage of this method is that existing shortest-path routing protocols can be employed without modifications. However, transmitting routing control packets at low power levels may reduce network connectivity and may also reduce the potential number of paths between the sources and destinations. If the source node is allowed to cache multiple paths for the same destination, it would not initiate another route discovery unless all the cached paths are broken. Therefore, fewer potential number of paths being cached at the source means more frequent route discoveries and higher routing overhead.

For the link-cost method, each link is assigned a cost and the blind flooding route discovery process is modified to discover the minimal cost path. The link cost should be defined such that short distance hops are preferred over long distance hops. For example, it could be a function of the sender-receiver distance.

6.2.3 Link-Cost Mini-hop Routing

To avoid substantial routing overhead, we combine the two approaches starting with the link-cost method.

The first question is how to define link cost. Since the goal of MHR is to discover paths that consume the minimal spatial resource, the link cost should reflect the spatial usage of a transmission over the link. According to Chapter 3, the spatial usage of an individual transmission is proportional to R_{cs}^2, where R_{cs} is the radius of the CSR and is determined by the transmit power P_t. However, when the MAC layer employs a power control scheme, we cannot predict the actual value of P_t and R_{cs}. As an alternative, we use d^2 to indicate the spatial usage of a transmission over the link, where d is the sender-receiver distance. This is because d determines P_{min}, which is a lowerbound on P_t, and, when transmitting at P_{min}, the spatial usage R_{cs}^2 is linearly proportional to d^2.

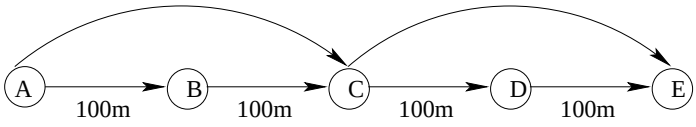


Figure 6.2: Duplicate RREQs must be broadcast to discover the minimal cost path.

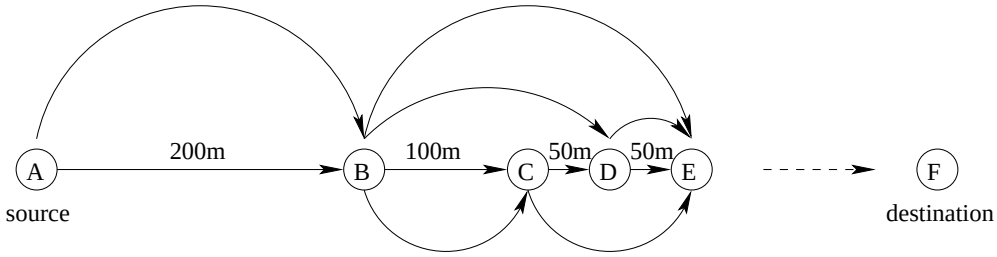


Figure 6.3: Controlling the RREQ broadcast sequence.

The second question is how to incorporate the link cost into the blind flooding route discovery process. To do this, duplicate RREQs must be broadcast as long as they carry paths that have lower cost than those carried by previous RREQs. We explain this process using Figure 6.2, where node A is the source and E is the destination. The transmit power for the routing control packets is P_{max} thus the transmit range $250m$. Each link is assigned a cost of d^2. Obviously, the minimal cost path from A to E is "$A \rightarrow B \rightarrow C \rightarrow D \rightarrow E$". When A broadcasts an RREQ, both B and C receive it. If node C gains the channel first and broadcasts the RREQ, both D and E receive it. Node E then replies to A with a RREP carrying path "$A \rightarrow C \rightarrow E$". Later, node B obtains the channel and broadcasts its RREQ. Although this RREQ is a duplicate for node C, node C has to rebroadcast it since it carries a path, "$A \rightarrow B \rightarrow C$" that has lower cost than "$A \rightarrow C$". The same action is taken at node D. Finally, node E replies to A with another RREP carrying the minimal cost path "$A \rightarrow B \rightarrow C \rightarrow D \rightarrow E$".

6.2.4 Controlling the Broadcast Sequence

Forwarding duplicate RREQs may result in considerable routing overhead, especially when node density is high. Figure 6.3 illustrates the scenario, where node A is the source and F is the destination. Node A initiates the route discovery process by broadcasting an RREQ for node F. We investigate the routing overhead by observing node E's behavior. Node E can receive RREQs from node B, C and D and may forward these RREQs depending on the path costs they carry. If node E happens to receive the RREQs as the following sequence, it has to forward all of them:

1. An RREQ from node B carrying path "$A \rightarrow B \rightarrow E$" with cost 80000.
2. An RREQ from node D carrying path "$A \rightarrow B \rightarrow D \rightarrow E$" with cost 65000.
3. An RREQ from node C carrying path "$A \rightarrow B \rightarrow C \rightarrow E$" with cost 60000.
4. An RREQ from node D carrying path "$A \rightarrow B \rightarrow C \rightarrow D \rightarrow E$" with cost 55000.

Here, each RREQ carries a path that has a lower cost than those carried by previous RREQs. Node E has to broadcast all four RREQs and the routing overhead is four times that in a normal blind flooding process.

To reduce these overheads, a plausible solution for a node is to wait until it receives all the potential RREQs and then broadcast the one that carries the minimal cost path. However, it is desirable to forward the RREQs as quickly as possible so that the path between the source and the destination can be established quickly. To solve this problem, the node forwards the first arriving RREQ immediately and stores the subsequent duplicates until it receives all the potential RREQs and then chooses the minimal cost one to broadcast. In this way, the node broadcasts RREQs at most twice during an individual route discovery process.

However, it is difficult to determine whether or not the node has received all the potential RREQs. As an alternative, the node waits for a duration, T, within which the minimal cost RREQ is likely to be received. A reasonable guess is that this duration is proportional to the distance from the node to the source, because the minimal cost RREQ likely comes from the nodes that are closer to the source. Fortunately, the first RREQ arrives at the node carries the least number of hops, n, from the node to the source, which is a rough indicator of the distance to the source. For the nodes that are in the same hop range, they could distinguish their distances by measuring the receive power level, P_r, of the first arriving RREQ. The closer the node to the sender, the higher the P_r. Thus, T is set to

$$T = (n + \frac{\mathcal{T}_{rx}}{P_r}) \times \epsilon \ , \tag{6.1}$$

where ϵ is a small constant and $\mathcal{T}_{rx}$ is the receiving threshold. As the fresh RREQ propagates from the source outward to the whole network, the nodes start waiting one after another. The further the node to the source, the longer T it waits and the later it broadcasts the minimal cost RREQ.

Consider Figure 6.3 again. After the source node A broadcasts the RREQ, node B forwards it immediately. Node C, D and E receive this fresh request from node B and determine that they are two hops away from the source. Because the three nodes are in the same hop range, they measure their receive power levels, P_r, of the RREQ to distinguish their distances more accurately. The closer the node to sender B, the higher the P_r. Thus, the waiting period at node E is longer than that at both C and D. Instead of rebroadcasting each duplicate RREQ immediately, node E waits until the end of the period, at which moment it should have received the RREQs from B, C and D, and node E rebroadcasts the one carrying the minimal cost path, "$A \rightarrow B \rightarrow C \rightarrow D \rightarrow E$". Thus, the number of transmitted RREQs are reduced substantially.

6.2.5 Route Reply

The RREQs are propagated throughout the whole network until they arrive at the destination or an intermediate node that knows the route to the destination.

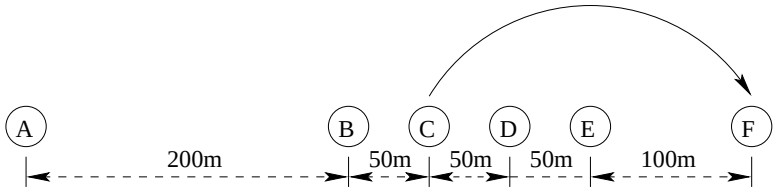
In DSR, there is an option that specifies whether an intermediate node is allowed to send a reply to the source. In NS2's [5, 6] implementation, it is enabled by default. Thus, we enable this option in MHR as well. The exception is that, if the intermediate node knows multiple routes to the destination, it replies with the route that has the minimal spatial usage cost, while, in DSR, it replies with the one that has the least number of hops.

There is another option that specifies whether the destination responses only to the first arriving RREQ or to all RREQs it receives. In our simulations, we allow multiple replies, because it improves network performance in all cases.

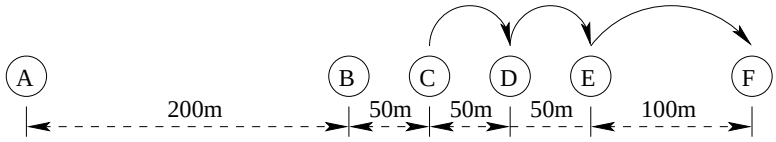
6.2.6 Excluding Long Distance Hops

We notice that several factors degrade the performance of the scheme above. First, to reduce routing overhead, many wireless routing protocols allow the source node to cache multiple paths to the same destination. When one path is broken, the source node resorts to another without initiating another route discovery process. In the scheme above, the source node receives not only the mini-hop paths but also the shortest-paths due to the propagation of both the minimal cost RREQ and the fresh RREQ. Once the mini-hop paths are broken, the source node would delivery subsequent packets along the shortest-path.

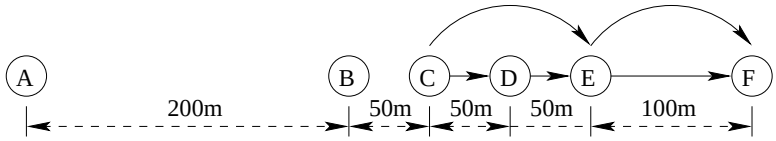
Second, if nodes are allowed to eavesdrop on routing information, they are likely to obtain long-distance hop information. This is because, when power control is employed at the MAC layer, the transmit power levels at long-distance hops are higher than those at short-distance hops and, therefore, long-distance hop information could be eavesdropped by a greater number of nodes.



(a) Identifying P_c on the per-network basis



(b) Identifying P_c on the per-node basis



(c) Identifying P_c on the per-connection basis

Figure 6.4: Identifying P_c.

Third, if an intermediate node knows only the shortest-path to the intended destination, it will propagate that information to the source as well as the nodes along the path.

A solution to these problems is to employ power control in the route discovery process. By transmitting routing control packets at reduced power levels, long distance hops can be excluded completely. Fortunately, after the initial link-cost route discovery process, the mini-hop paths reveal connectivity between the source and the destination and, therefore, P_c can be identified.

6.2.7 Identifying P_c

The critical power level, P_c, at which routing control messages are transmitted, can be identified on a per-network, per-node or per-connection basis.

For the per-network case [113], the whole network maintains a common P_c that guarantees network connectivity. All routing control packets are transmitted at P_c no matter what the source-destination pair is. The per-network based power control is not efficient, especially when clusters exist. For example, in Figure 6.4(a), to prevent node A from being isolated, P_c has to be high enough to cover a distance of $200m$. Assume that node C is the source and node F is the destination. When node C broadcasts a RREQ at P_c, node F receives it and replies immediately with the path "$C \rightarrow F$", which is not the minimal cost one.

For the per-node case [112, 115], each node maintains a P_c that is high enough to reach the furthest node in its connectivity set and it uses this power level to transmit routing control packets. The advantage of per-node based power control is that each node determines the hop distance according to its connectivity to the network. In Figure 6.4(b), each node's connectivity set consists of the immediate neighbor on the right and left. Therefore, both node C and node D set their P_c to the power level whose transmit range is $50m$, and node E set its P_c to the power level whose transmit range is $100m$. When node C initiates the route discovery process for node F, each node uses its own P_c to forward the RREQ and the minimal cost path "$C \rightarrow D \rightarrow E \rightarrow F$" can be discovered directly. However, to maintain the connectivity set, considerable overhead is needed.

To avoid the inefficiency of the per-network case and avoid the maintenance of connectivity sets in the per-node case, we employ per-connection based power control in MHR. That is, each source-destination pair maintains a P_c which guarantees connectivity between the source and the destination. The source-destination pair identifies its P_c as follows. After an initial link-cost route discovery, the source obtains the minimal cost path, which reveals the connectivity between the source and the destination. The source node identifies the longest hop distance, h_{max}, in this path and sets P_c to be the power level whose transmit range is slightly greater

than h_{max}. In the next route discovery process for the same source-destination pair, routing control packets are transmitted at the P_c. Thus, the hop distances would not exceed h_{max}. If, for some reason such as a dramatic change of node connectivity, the route discovery fails, the source node resets P_c to P_{max} and initiates another route discovery process immediately. Of course, the side effect is that routing overhead increases.

Although the per-connection based power control may not be able to discover the minimal cost path, it can when combining with the link-cost routing scheme. In Figure 6.4(c), after an initial link-cost route discovery for the source-destination pair "C-F", node C obtains the mini-hop path, "$C \rightarrow D \rightarrow E \rightarrow F$", and identifies P_c to be the power level whose transmit range is $100m$. Next time when node C initiates a route discovery process for node F, every node forward the RREQ at the P_c. The path, "$C \rightarrow E \rightarrow F$", will be discovered. Meanwhile, duplicate RREQs are forwarded to follow the link-cost routing scheme. Thus, the source node C receives not only "$C \rightarrow E \rightarrow F$" but also "$C \rightarrow D \rightarrow E \rightarrow F$".

6.2.8 Node Mobility

The scheme above works well in low mobility networks. Node mobility impacts the performance of our scheme in two ways. First, transmitting routing control packets at P_c, which is determined by previously discovered mini-hop paths, may not be able to successfully discover paths in subsequent route discover processes due to dramatic changes in node connectivity. Second, path costs vary as nodes move, which impacts path selection at the source node.

The first problem is not a big issue when node mobility is low and traffic load is high. In such case, transmissions are likely to fail due to intense channel contention, resulting in frequent route outages and thus frequent updates on P_c. Since node mobility is low, network connectivity would not change dramatically

between consecutive route discovery processes.

To address the second problem, the source node may update the path cost while delivering data packets. That is, every forwarding node along the path maintains the most recent path cost from itself to the destination. A downstream node feeds back the most recent cost to its upstream node using MAC layer ACK packets. The upstream node then updates its own cost accordingly and propagates this information backward until the source node is updated.

However, we did not implement this scheme in the version of MHR discussed below. There are two reasons. First, our subsequent simulations show that MHR performs well in low mobility networks. Second, the characteristics of typical multihop wireless networks are unclear at present and we expect that low mobility networks with heavy traffic are not uncommon in real world. For example, in a conference, people move at walking speed occasionally and heavy traffic is loaded onto the network due to information exchanging and internet surfing. Thus, we leave the mobility issue as a future research topic.

6.2.9 MHR and Rate Adaptation

In this Chapter, we focus on the synergy between MHR and power control. Another interesting question is what is the synergy between MHR and rate adaptation. Reducing hop-distance is beneficial for increasing transmission rate at each hop. But, it comes at the cost of a greater number of hops. We briefly analyze the combined effect below.

Assume that the spatial usage of delivering a packet along the shortest path from a source to a destination is Ω_0 and that the average hop distance is d and the maximum transmission rate at each hop is r. If we use a mini-hop path to replace the shortest path such that each original long hop is replaced by k short hops, the short hop distances would be $\frac{d}{k}$ and the maximum transmission rate would be

$r + \alpha W_c \ log_2(k)$ according to Shannon's Theorem. Meanwhile, the number of hops in the mini-hop path is k times greater than that in the shortest path. Thus, the spatial usage of delivering a packet along the mini-hop path is $\frac{kr}{r+\alpha W_c \ log_2(k)}\Omega_0$, which increases asymptotically as k increases. This suggests that network performance may not be able to benefit from the synergistic operation of rate adaptation and mini-hop routing.

6.3 Details of the MHR Protocol

Based on these considerations, we present the details of MHR. MHR is an on-demand mini-hop routing protocol that combines the link-cost and the power-control methods. The basic idea is that each link is assigned a spatial usage cost and the route discovery process attempts to find the minimal cost path. To exclude long distance hops, the source node learns connectivity information to the destination from previous discovered mini-hop paths and tunes the transmit power of routing control packets in the subsequent route discovery process.

6.3.1 Link Cost

To reflect the idea of spatial usage efficiency, we define the link cost as d^2, where d is the hop (or link) distance. A path cost is the summary of all the link costs along the path. As we stated above, even though d^2 may not be the actual spatial usage of the transmission along the hop, it is linearly proportional to the actual spatial usage if the sender transmits at P_{min}.

6.3.2 The Route Discovery Process

Before describing the route discovery process in detail, we present the packet format modifications and additional data structures needed in each node.

As in DSR, the RREQ packet contains the source address, *sid*, the destination address, *did* and a sequence number, *seq*, which indicate the freshness of the request. Additionally, we add three more fields:

1. A route record that contains not only the sequence of the hops, as DSR does, but also each hop-distance traversed by the RREQ.

2. A path cost field, "PC", which is the sum of the link costs along the path traversed by the RREQ. This field is redundant, because it can be calculated according to the hop distances in the route record. We add this field for efficiency purposes.

3. A power level field, "PL", which indicates the power level, P_c, at which the RREQ should be transmitted. This field is set by the source node.

To implement the idea of per-connection based power control, each source node maintains a Critical Power (CPR) table, in which each entry contains two fields:

1. The destination address, *did*.

2. The critical power, P_c. During the route discovery process for the destination *did*, routing control control packets are transmitted at P_c. Initially, P_c is set to P_{max}. The source node updates P_c every time a route discovery process for the destination is completed.

To implement the idea of controlling RREQ broadcast sequence, a node maintains a rebroadcast timer, (RBTimer), for each source-destination pair. When the node receives a fresh RREQ for a specific source-destination pair, it triggers the RBTimer and sets the expiration time T according to (6.1). For duplicate RREQs, the node simply stores them. At the moment the RBTimer fires, the node

chooses the minimal cost one to broadcast. To store these information, we modify the Route Request Table (RRT) in DSR to be the Minimal Cost Table (MCT). The RRT records the most recent route request for each source-destination pair and, therefore, the node can determine whether an RREQ is fresh or a duplicate. The entry in the RRT contains three fields, *sid*, *did* and the most recent route request sequence number, seq_r. In addition to these fields, entries in the MCT contain the following fields:

1. The RBTimer. It is set every time the node receives a fresh request for the *sid-did* pair.

2. The minimal path cost, PC_{min}, from the node to *sid*, which was discovered in the most recent route discovery process. Every time the node receives a fresh RREQ for the *sid-did* pair, this field is reset. When the node receives a duplicate RREQ, it compares PC_{min} with the "PC" field in the RREQ and updates PC_{min} accordingly.

3. The RREQ packet carrying the path with cost PC_{min}. This RREQ will be broadcast when the RBTimer fires.

Before the source node S broadcasts an RREQ for the destination node D with sequence number s, it looks in the CPR table for the corresponding P_c. Thereafter, the source node sets the "PC" field to 0 and sets the "PL" field to P_c in the RREQ header and transmits the RREQ at P_c.

Suppose that node i receives a RREQ from node j. It measures the receiving power level, P_r, estimates the distance to node j, d_{ij}, adds its ID as well as d_{ij} to the route record, and updates the "PC" field in the RREQ such that $PC = PC + d_{ij}^2$. Thereafter, node i looks for the entry corresponding to triple $< S, D, s >$ in the MCT table, determines whether the RREQ is fresh or a duplicate, and updates the entry as follows:

- If the RREQ is fresh, it counts the number of hops, n, in the route record and starts the RBTimer with expiration time calculated by (6.1). Meanwhile, it sets $PC_{min} = PC$ and rebroadcasts the RREQ at P_c immediately.

- If the RREQ is a duplicate, node i compares PC_{min} with the "PC" field in the RREQ. If $PC > PC_{min}$, node i discards the RREQ. Otherwise, it sets $PC_{min} = PC$ and saves this RREQ.

When the RBTimer fires, node i broadcasts the corresponding RREQ, which is the minimal cost one among these received during the waiting period.

The RREQ is propagated throughout the whole network until it arrives at the destination, which then copies the route record and the "PC" information into an RREP and sends the RREP back to the source. If an intermediate node, that knows the route to the destination, receives the RREQ, it appends the route in its cache to the route record, updates "PC" to be the sum of the two routes, copies the route record and the "PC" information into an RREP and sends the RREP back to the source. The RREP is forwarded backward along the route traversed by the RREQ. Upon receiving the RREP, the source node retrieves the route record and "PC" and stores them into its route cache.

6.3.3 Route Selection

A source may be allowed to catch multiple path to the same destination. It uses two criteria to decide along which route data packets are delivered: the path cost and the number of hops. The source node always chooses the route that has the minimum spatial usage cost. However, if two routes have the same cost, the source node uses the number of hops to break the tie.

6.3.4 Calculating P_c

As we stated in Section 6.2, P_c is determined by the hop distances of the paths discovered in the previous route discovery process.

Assume that the source receives n different paths to the same destination, denoted as p_i where $i = 1, 2, ..., n$. For each path p_i, the source node finds the maximum hop distance h^i_{max}. Note that the route record in the RREP contains not only the hop sequence of the path but also each hop distance. The source node identifies the minimum of these maximal hop distances, e.g. $h^{min}_{max} = min\{h^1_{max}, ..., h^i_{max}, ..., h^n_{max}\}$. P_c is slightly higher than the power level whose transmit range is h^{min}_{max}. Thereafter, P_c is used in the next route discovery process for the same destination. If route discovery fails at power level P_c, the source node resets P_c to P_{max} and initiates another route discovery immediately.

6.3.5 MHR and Distance Estimation

In MHR, the link cost is set to d^2, where d is the sender-receiver distance of the link. Again, if each wireless node is capable of positioning (Section 2.10), d can be readily obtained.

If the wireless nodes do not have a positioning system, MHR can acquire the data packet transmit power, P_t, that would be used for this link from the MAC layer, and use a function of P_t, $\mathcal{F}(P_t)$, as the link cost. $\mathcal{F}(P_t)$ is the size of the CSR corresponding to P_t and it is defined as

$$\mathcal{F}(P_t) = \left(\frac{\beta P_t}{\mathcal{T}_{cs}}\right)^{\frac{2}{\alpha}} . \tag{6.2}$$

Here, we assume that the path-loss exponent, α, is known.

P_t is predicted during the blind flooding process when a node receives an RREQ from a sender. The node estimates the receive power, P_r, of the RREQ and predicts the P_t at this link. To predict P_t, the receiver needs to estimate the channel

gain using the SINR estimation techniques introduced in Section 2.8. Notice that P_t may not be the actual power that will be used to transmit data packets since it is affected by future interference conditions. Just as d^2, $\mathcal{F}(P_t)$ is a rough indiction of the spatial utilization of the transmissions at this link.

If the receiver is not capable of estimating SINR, it can estimate the channel condition by approximating P_r as the total measured power, as discussed in Section 2.9. Using the lower bounds in (2.13) or (2.14), the receiver may overestimate the spatial utilization, while, using the upper bounds, the receiver may underestimate the spatial utilization. Because route selection is based on the relative spatial usage of each route, it is unknown what the impact would be on MHR if using the power approximation. We leave this issue as a potential future research topic.

6.4 Simulation Results

We implemented MHR and evaluated its performance in NS2 [5, 6]. MHR alone can not improve network performance. Instead, it has to work with a MAC layer power control scheme. By delivering packets along short hops, MHR makes it possible for the MAC layer power control scheme to reduce the transmit power even lower and, therefore, to improve network performance even further. To understand this synergistic cooperation, we evaluate the performance of three strategies: shortest-path routing (SPR) without MAC layer power control (SPR_NPC), SPR with MAC layer power control (SPR_PC), and MHR with MAC layer power control (MHR_PC). By comparing the performance of these strategies, we will have a clear understanding of how much benefit MHR and the MAC power control scheme would bring about on network performance. We use DSR [35] as the shortest-path routing protocol and the IEEE 802.11 DCF as the MAC protocol without power control.

In Chapter 4, we proposed a MAC power control scheme, OTP. Our initial simulations showed that MHR combining with OTP did not improve network

performance. This is because MHR reduces hop distances at the cost of a greater number of hops. Although short hop distance reduces spatial usage of individual transmissions, a greater number of hops increases the spatial usage of delivering an individual packet from source to destination. The combined effect depends on how much impact each individual factor would have on the network spatial utilization. If the MAC layer power control scheme reduces transmit power in an overly conservative manner, such as OTP, the spatial resource conserved by using the short distance hops would be eliminated by transmitting the packet along the greater number of hops. For this reason, we proposed EOTP, which adjusts transmit power more aggressively than OTP. In the subsequent simulations, we use EOTP as the MAC layer power control scheme.

Network performance is evaluated based on network throughput, packet delivery latency, and routing overhead. The packet delivery latency is the time from when a data packet is generated at a source to the time when it is received at a destination. Routing overhead is the total number of bytes in the routing control packets that have been transmitted divided by the number of data packets received at the destinations. Therefore, it is a per data packet overhead.

As Muqattash *et al.* did in [115], we consider two types of topologies, uniform and clustered, to test two different traffic patterns. In the unform topology, traffic flows are randomly distributed, while in the clustered topology, traffic flows demonstrate locality. Additionally, we investigate the impact of network size and node mobility on the performance of MHR.

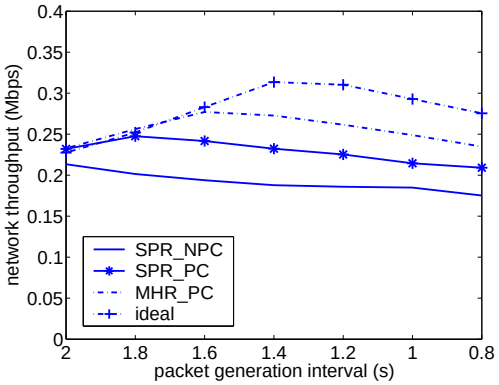
6.4.1 Uniform Topologies

We evaluate the performance of MHR in networks with nodes that are uniformly distributed. We use a 1100m by 900m grid to represent the uniform topology, where adjacent nodes are 100m apart. Let P_{con} be the power level whose transmit

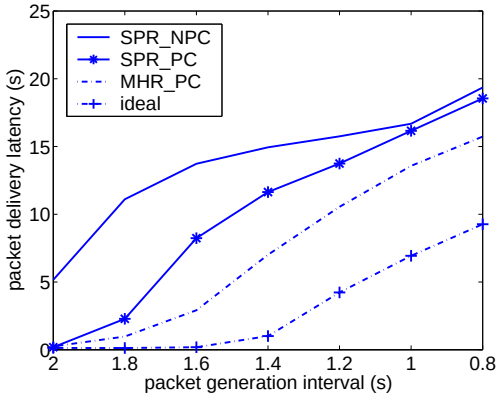
range is barely above 100m. In this topology, it is the lowest power level that guarantees network connectivity. If we employ DSR and 802.11 and transmit all packets, including routing control packets and data packets, at P_{con}, we minimize the hop distance and the transmit power of each individual transmission without additional modifications to these protocols. Thus, it should achieve the highest network throughput. This strategy is denoted as "ideal".

To fully utilize the network spatial resource and evaluate the protocol performance fairly, traffic flows are distributed throughout the whole network. To do so, each node sends CBR traffic to a randomly chosen destination and this source-destination relationship is maintained till the end of the simulation. Each traffic flow starts at a time randomly chosen within 50 seconds. All packet lengths are 512 bytes. We measure network performance as a function of traffic load.

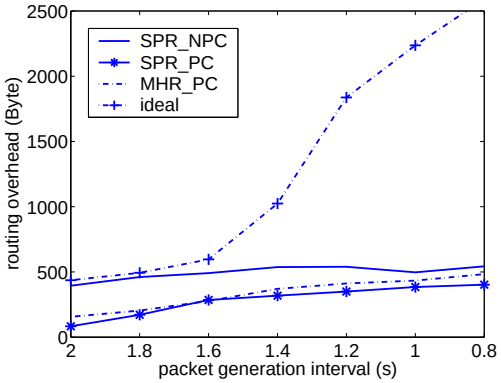
Figure 6.5(a) shows the network throughput. The x-axis is the packet generation interval at the source nodes and the smaller the interval, the higher the traffic load. The y-axis is the network throughput in Mbps. When the traffic load is low, the four strategies have the same throughput because the network spatial resource is sufficient for all transmissions and power control is not needed. As the traffic load increases, the demand for spatial resource increases. SPR_PC outperforms SPR_NPC because it employs power control at the MAC layer. But, SPR restricts the effect of the MAC layer power control schemes by delivering packets along long hops. Unlike SPR, MHR delivers packets along short hops so that each individual transmission can reduce its power to a lower level. Therefore, MHR_PC outperforms SPR_PC. Because the "ideal" strategy minimizes both the hop distances and the transmit power of each individual transmission, it achieves the highest network throughput. MHR_PC underperforms "ideal" because MHR does not guarantee that all the data packets are delivered along the minimal cost path. For example, it allows the source nodes to cache both shortest-paths and mini-hop paths. Once the mini-hop paths



(a) network throughput



(b) packet delivery latency



(c) routing overhead per data packet

Figure 6.5: Performance of the four strategies in the uniform topology.

are broken, the source nodes would deliver data packets along the shortest-paths.

Figure 6.5(b) shows the average packet delivery latency. Packet delivery latency consists of two major parts: forwarding latency and queuing latency. The forwarding latency is determined by the number of hops the packet is forwarded along, while the queuing latency is determined primarily by the channel contention in wireless networks. When the traffic load is low, queuing delay is minimal and forwarding latency dominates the total packet delivery latency. However, when the traffic load is high, forwarding latency is less significant and queuing latency is dominant. SPC_PC has lower packet delivery latency than SPC_NPC, because the MAC layer power control scheme reduces channel contention and queuing latency by reducing the transmit power of each individual transmission. Moreover, MHR makes it possible for the MAC layer power control scheme to alleviate channel contention and the queuing latency in a further step. Although it comes at the cost of a greater number of hops, which results in higher forwarding latency, the combined effect is that MHR_PC achieves lower packet delivery latency than SPR_PC because the forwarding latency is less significant than queuing latency when the traffic load is high. Because the "ideal" strategy minimizes the hop distances and the transmit power of each individual transmission, it minimizes the channel contention. Therefore, it performs best among the four strategies.

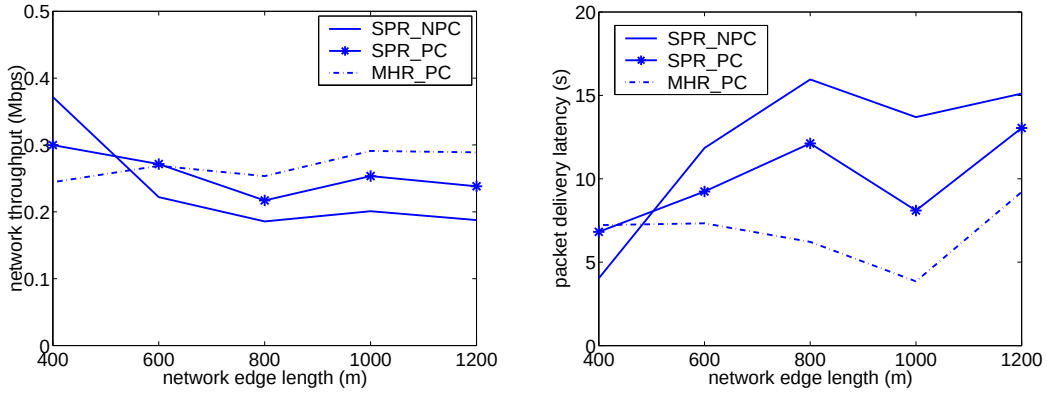
Figure 6.5(c) shows the routing overhead. The x-axis is the packet generation interval and the y-axis is the per-packet routing overhead in bytes. Although the nodes are stationary, route outages occur frequently due to channel contention. All four strategies show increasing routing overhead as the traffic load increases. SPC_PC has lower routing overhead than SPC_NPC. This is because the MAC layer power control scheme reduces channel contention, resulting in the established paths being more stable and the route discoveries being less frequent. Also, the simulation shows that MHR_PC brings about slightly higher routing overhead than

SPC_PC. This is caused by the combined effect of several factors. On one hand, MHR_PC reduces channel contention and stabilizes the paths even further, which results in less frequent route discoveries than SPC_PC. In addition, MHR increases the number of successfully delivered packets and therefore reduces the per-packet routing overhead. On the other hand, MHR allows both the fresh and duplicate RREQs to be forwarded, which brings about a greater number of routing control packets and higher routing overhead per discovery. Moreover, the size of each routing control packet in MHR is larger than that in DSR. Therefore, the overall effect is that MHR_PC has a slightly higher routing overhead than SPC_PC.

However, we observe that the "ideal" strategy generates enormous routing overhead when the traffic load is high. This is because "ideal" uses the minimal power level, P_{con}, to control the route discovery process, which barely maintains the network connectivity. Thus, the number of discovered paths is minimized. In our simulation, we allow the source node to cache multiple paths for the same destination. The source node would not initiate another route discover process unless all the cached paths are broken. Fewer number of discovered paths result in more frequent route discoveries and higher routing overhead. Unlike "ideal", MHR identifies P_c based on the maximal hop distance of the mini-hop paths, which leads to a stronger network connectivity. Therefore, MHR has less frequent route discoveries and lower routing overhead than "ideal".

6.4.2 Network Size

Chapter 3 points out that mini-hop routing and MAC layer power control scheme improve network performance only when the network size is relatively large. We perform simulations to demonstrate the impact of network size on the performance of MHR_PC by using a grid topology and varying the network size from 400m by 400m, 600m by 600m to 1200m by 1200m. The neighboring nodes are 100m apart.



(a) network throughput

(b) packet delivery latency

Figure 6.6: The impact of network size on the performance of MHR.

Every node randomly chooses a destination and sends CBR traffic until the end of the simulation. The CBR traffic rate is fixed under each network size and is set to be high enough such that all strategies could achieve their achievable throughput.

Figure 6.6(a) shows the network throughput. The x-axis is the edge length of the grid topology and the y-axis is the throughput in Mbps. When the network size is small, for example 400m by 400m, MHR_PC underperforms both SPR_NPC and SPR_PC. This is because, in small-sized networks, all transmissions are crowded into a small area. Most of the spatial resource conserved by the MAC power control scheme falls outside of the network and cannot be utilized by other transmissions. MHR attempts to improve the performance of the MAC layer power control scheme by reducing hop distances at the cost of a greater number of hops. But, in our simulation, the minimum hop distance is 100m, which is not short enough to let the MAC power control scheme decouple transmissions in a 400m by 400m network area. In addition, the greater number of hops leads to more transmissions, more intense channel contention and lower network throughput. However, as the network size increases, transmissions are more widely spread and the conserved spatial resource can be utilized by the transmissions that are far apart. Under this circumstance,

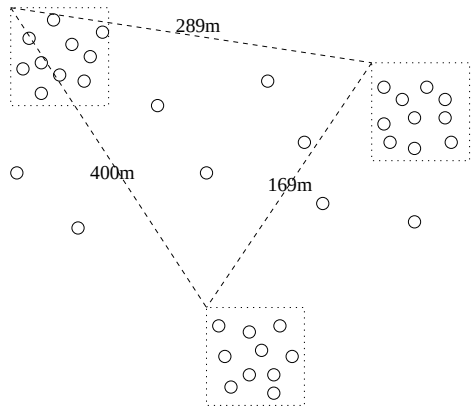


Figure 6.7: A clustered topology

short hop distances do help the MAC layer power control scheme decouple transmissions. Although it comes at the cost of a greater number of hops, the overall effect improves network throughput. Therefore, when network size increases, MHR_PC outperforms both SPR_NPC and SPR_PC.

Figure 6.6(b) shows the packet delivery latency. Again, when the network size is small, MHR_PC underperforms both SPR_NPC and SPR_PC. For the same reasons stated above, the MAC layer power control scheme cannot alleviate the channel contention and queuing latency. Since MHR forwards data packets along a greater number of hops, it increases not only the forwarding latency but also the queuing latency due to the exacerbated channel contention. However, when the network size increases, the MAC layer power control scheme is able to decouple transmissions and alleviate channel contention as well as queuing latency. This effect is further improved with the help of MHR. Although MHR forwards data packets with a greater number of hops, which increases the forwarding latency, the combined effect improves packet delivery latency.

6.4.3 Clustered Topologies

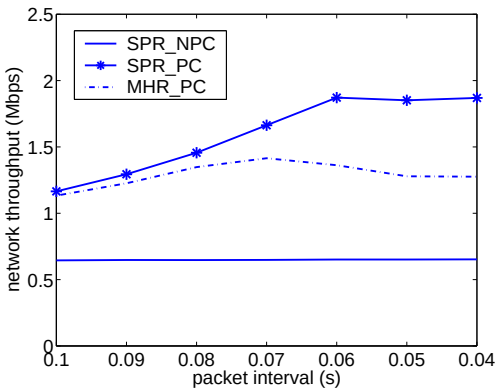
In general, in real world, nodes may group together and form clusters. For example, in a big conference hall, each room may be a cluster of nodes. In the clustered topologies, traffic flows may demonstrate locality [114, 115]. That is, most traffic flows are within the clusters. We evaluate the performance of MHR using networks with these characteristics.

Figure 6.7 shows a clustered topology, where three 100m by 100m clusters are in a 600m by 600m network area. Each cluster contains 10 randomly distributed nodes. Every node sends CBR traffic to a randomly chosen destination within the same cluster. The distance between these clusters are 289m, 169m and 400m respectively. These positions are arranged such that the three clusters could decouple from each other when power control is employed at the MAC layer. We measure the network performance with varying traffic load.

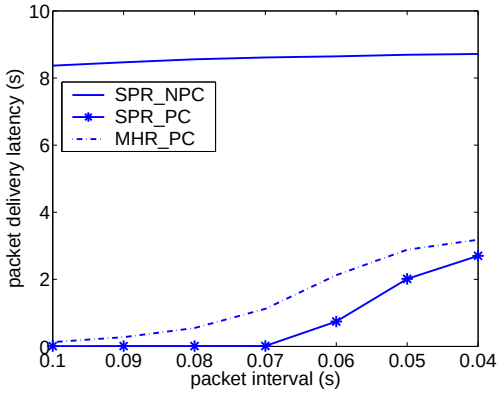
Figure 6.8(a) shows the network throughput. The x-axis is the packet generation interval at the source nodes and the y-axis is the network throughput in Mbps. SPR_PC outperforms SPR_NPC because the MAC layer power control scheme decouples the three clusters and, therefore, the transmissions in different clusters can transmit concurrently. However, because the size of each cluster is small, MHR degrades the performance within each cluster, and we observe that MHR_PC underperforms SPC_PC.

Figure 6.8(b) shows the packet delivery latency. SPR_PC outperforms SPR_NPC, because the MAC power control scheme decouples the three clusters and alleviates the channel contention, which leads to lower queuing delay and packet delivery latency. Also, because the size of each cluster is small, MHR increases the channel contention and the queuing latency within each cluster, and we observe that MHR_PC underperforms SPR_PC.

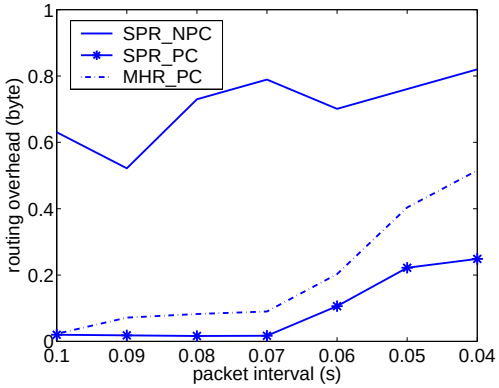
Figure 6.8(c) shows the routing overhead. The y-axis is the per-packet rout-



(a) network throughput



(b) packet delivery latency



(c) routing overhead per data packet

Figure 6.8: Performance of the three strategies in the clustered topology.

ing overhead in bytes. SPR_PC outperforms SPR_NPC because of two reasons. First, the MAC layer power control alleviates the channel contention and reduces route outages. Second, the network throughput in SPC_PC is much higher than that in SPC_NPC. Therefore, the per data packet routing overhead is lower. Moreover, MHR_PC underperforms SPR_PC because MHR increases channel contention within each small-sized cluster.

6.4.4 Node Mobility

Finally, we evaluate the impact of mobility on the performance of MHR. In our simulation, nodes move according to the random waypoint mobility model [6]. That is, each node independently chooses a random starting point and waits there for a duration called the pause time. It then randomly chooses a destination, and moves there at a speed chosen uniformly between 0 and a maximum velocity, v_{max}. When the node arrives at the destination, it again waits for the pause time, and then begins moving at a new randomly chosen velocity to a new randomly chosen destination. each node independently repeats this movement pattern throughout the simulation.

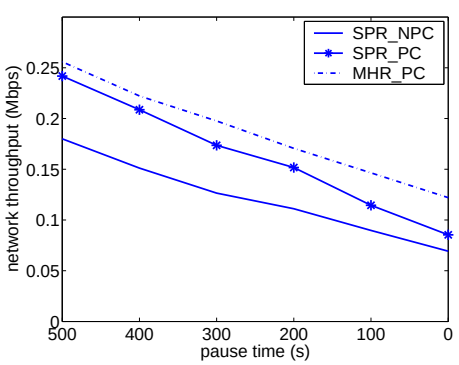
We randomly put 120 nodes in a 1000m by 1000m area. Each node sends CBR traffic to a randomly chosen destination at the rate of 0.525 packet/second (or 1.6 seconds/packet). All packets are of length 512 bytes and the simulation lasts 500 seconds. v_{max} is set to $3m/s$ to imitate the slow moving vehicle speed. Because the pause time indicates the degree of constant motion, we measure the network performance with varying pause time.

Figure 6.9 (a) shows the network throughput, where the x-axis is the pause time and the smaller the pause time, the higher the mobility. When mobility increases, the frequency of route outage increases, which results in the performance degradation for all three strategies. For the same reason stated for the uniform topology simulation, MHR_PC outperforms both SPR_PC and SPR_NPC because

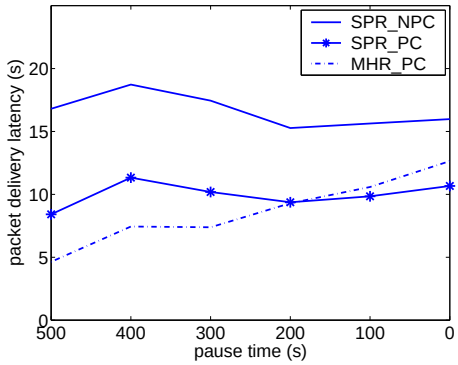
of employing MAC layer power control and mini-hop routing.

Figure 6.9(c) shows the routing overhead per successfully delivered data packet. The x-axis is the pause time, and the y-axis is the routing overhead in bytes. We had expected that, when node mobility increases, route outages in MHR would occur much more frequently than in SPR because mini-hop paths contain a greater number of hops than shortest-paths and they are more likely to be broken. But the simulation shows that, when node mobility increases, route outages in MHR occur much less frequently than in SPR. We investigated the simulation and found that, with EOTP, mini-hop paths are more resistance to breaking due to node mobility than shortest-paths. This is because, a mini-hop path consists of a greater number of short hops and each hop can extend to the maximal hop distance. In another words, the mini-hop path could potentially span a longer distance than the shortest-path could. When the source and destination are moving far away from each other, the mini-hop path is more sustainable. Moreover, during each route discovery process, MHR discovers not only mini-hop paths but also shortest-paths. The source node can cache more route information. When one path is broken, the source node would resort to another path without initiating a route discovery. Therefore, we observe that MHR_PC brings about less routing overhead than SPR_PC and SPR_NPC.

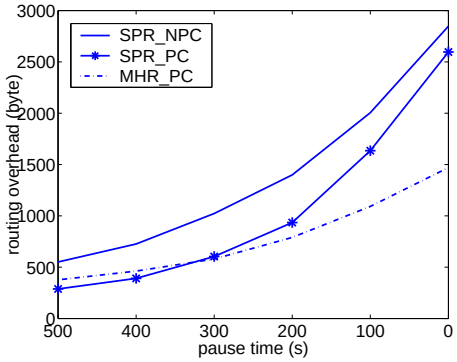
The frequent route discovery in SPR brings about an unexpected effect on packet delivery latency, as shown in Figure 6.9(b). According to the blind flooding technique, the discovered path has the minimum delivery latency. But, as nodes move and traffic pattern varies, the path may not be able to maintain the minimum delivery latency. Frequent route discovery allows SPR to update the path frequently and delivery packets along the current minimal latency paths. Obviously, the side effect of frequent route discovery in SPR is degraded network throughput and increased routing overhead. On the other hand, the paths in MHR are more



(a) network throughput



(b) packet delivery latency



(c) routing overhead per data packet

Figure 6.9: The impact of node mobility on the performance of three strategies.

sustainable and not very sensitive to node mobility or traffic pattern variations, which means these paths do not guarantee the minimal delivery latency. Although MHR_PC reduces channel contention and queuing latency, the combined effect is that, when node mobility increases, MHR_PC underperforms SPR_PC.

6.5 Impact of Link Loss Rate

Some routing protocols have been proposed that choose routes based on link loss rate [107, 108, 109] for stationary multihop wireless networks. Couto *et al.* [107]

defined the metric, *expected transmission count* (ETX). The ETX of a link is the predicted number of data transmissions required to send a packet over that link, including retransmissions. The ETX of a route is the sum of the ETX for each link in the route. Routes are selected such that they minimize the expected total number of packet transmissions (including retransmissions) required to successfully deliver a packet to the ultimate destination. Draves *et al.* [108] assigned link cost based on the *Expected Transmission Time* (ETT) of a packet over the link. ETT is a function of the link loss rate and the link bandwidth. The individual link weights are combined into a path metric called *Weighted Cumulative ETT* (WCETT). Routes are selected to minimize the WCETT between a source and a destination.

Incorporating these metrics into MHR is an interesting idea. In doing this, route selection would be based on both spatial utilization and link loss rate. We leave this idea as potential future work for two main reasons:

1. First, packet loss, or link loss, is caused by interference, small-scale fading, or the shadowing effect. Our work is based on the assumption that we ignore fast fading and the shadowing effect. Therefore, in our current research domain, the main reason that may cause packet loss is interference. The goal of Chapter 4 and 5 is to design MAC layer power control schemes to counter interference and prevent potential packet loss. In this sense, the MAC layer power control scheme eliminates the major causes of link-loss.

2. Second, the idea of link loss rate does not invalidate the design rationale of MHR. The impact would be such that, when we choose mini-hop routes, some links will be potentially excluded because of bad link status. Additionally, with MAC layer power control, link reliability can be improved.

If we were to incorporate the link-loss rate factor into MHR, the key issue is to find an appropriate way to measure either ETX or ETT. Both [107] and [108]

estimated the ETX or ETT metric by sending probe packets periodically, for example, every second, and estimate the probability of packet receptions for a predefined duration. These are nice simple schemes for stationary wireless networks where channel condition do not vary quickly. For mobile wireless networks, an efficient way to estimate link status might be to acquire channel information from the physical layer using cross-layer design. At the network layer, each source maintains multiple routes to the same destination and dynamically switching to the current most reliable route to deliver data packets. We leave this for future work.

Chapter 7

Load-Sensitive Routing

To achieve the maximum network throughput, the networks spatial resource needs to be fully utilized. Load-sensitive routing is a strategy that circumvents hot spots and routes traffic flows into under-utilized space.

We will show that the load-sensitive routing can potentially improve network performance in three ways.

First, load-sensitive routing could improve throughput. Consider the scenario in Figure 7.1. Two source-destination pairs, A-B and C-D, attempt to communicate at full channel capacity W. Node A delivers data packets to B along "A→E→B" and node C delivers data packets to D along "C→E→D". The two flows have to share a single channel capacity, W, because they are all forwarded by node E and the area around E is congested. If one of the flows could detect the congestion and bypass the hot spot, for example flow C-D takes the route "C→F→D", the two flows can be decoupled from each other. Thus, each flow is served by the full channel capacity, resulting in a higher throughput.

Second, load-sensitive routing could decrease end-to-end packet delivery latency. Packet delivery latency consists of two major parts: forwarding latency and queuing latency. In wireless networks, when the load is high, the queuing latency is

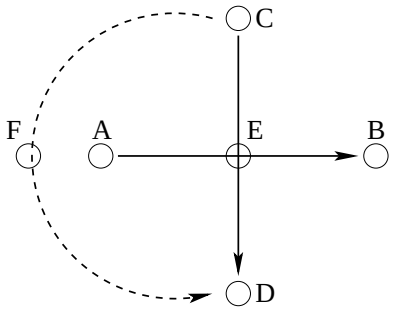


Figure 7.1: Load-sensitive routing.

much more significant than forwarding latency [74]. Delivering data packets through congested areas increases queuing delay substantially due to intense channel contention. Load-sensitive routing may be able to avoid such substantial queuing delay and, therefore, decrease the end-to-end delivery latency.

Third, load-sensitive routing could decrease routing overhead. This is because paths that go through congested areas are likely to be broken due to intense channel contention. By circumventing the hot spots, paths established by load-sensitive routing are much more stable, which would result in less frequent route outages and lower routing overhead.

Motivated by the above considerations, we present the design of several load-sensitive routing protocols that circumvent network hot spots and deliver packets along idle space.

7.1 Related Work

Several load-sensitive routing protocols have been proposed [98, 99, 100, 101, 102]. In [98], a node drops routing control messages if it detects that the load around it is heavy. Thus, the discovered paths do not contain congested nodes. The advantage of this strategy is that existing routing protocols can be employed with trivial modifications. But, the disadvantage is that some source nodes may not be able to

find paths to intended destinations because of heavy traffic. This is not appropriate in terms of fairness and preserving network connectivity.

Other protocols employ link cost routing. That is, each link is assigned a cost, reflecting the traffic load at the sender, and the routing protocols attempt to discover the minimal cost path. These protocols differ from each other in the definitions of load.

Hassanein *et al.* [99] proposed that a node counts the number of flows carried by itself and its neighbors. Since different flows have different actual load, this definition cannot properly indicate the node's congestion status.

Lee *et al.* [100] used the outgoing packet queue length at a node as the indication of the traffic load. However, in wireless networks, the node competes for the channel with those in the same carrier sense range. A node that has no packet in its sending buffer might be congested because of the heavy traffic around it. Wu *et al.* [101] improved [100] such that they defined the traffic load at the node to be the number of buffered packets at the node and its neighbors. Unfortunately, they ignored the fact that the nodes in the carrier sense range compete for the channel as well.

Song *et al.* [102] used the average packet delay at a node to indicate the traffic load. However, there is no way to estimate the average delay if the node does not forward packets. Imagine the node that does not carry any traffic flows and yet is congested by the heavy traffic of the neighbors. Although the average packet delay at the node is the minimum, it is not desirable to forward data packets through it.

7.2 Design Rationale

We base our design of the load-sensitive routing protocol on several considerations, including congestion detection, on-demand routing and the approaches of discovering non-congested paths.

7.2.1 Congestion Detection

A crucial issue in designing a load-sensitive routing protocol is accurate congestion detection. Existing load sensitive routing protocols [99, 100, 101, 102] use metrics such as the number of traffic flows, the number of buffered packets or packet delay to indicate traffic load and detect congestion. As we mentioned above, these metrics do not apply in some circumstances.

We claim that channel utilization, $chUtil$, is an appropriate metric to detect congestion. It is defined as follows. Assume that a node monitors the channel for a duration of T_m and, within this period, it detects the channel to be busy for a total duration of t_b. $chUtil$ is defined as

$$chUtil = \frac{t_b}{T_m} \ . \tag{7.1}$$

The node considers the channel to be busy if the node is transmitting, receiving or the node senses that the channel is reserved by other ongoing transmissions. With $chUtil$, the node can correctly detect congestion even if it does not carry any traffic. In addition, this metric captures the average congestion status over a ceratin interval and thus prevents the instantaneous congestion status from affecting route selection.

7.2.2 On-Demand Routing

Our load-sensitive routing protocol is based on on-demand routing for two main reasons.

First, on-demand routing is considered to be more appropriate than table-driven routing in multihop wireless networks [31], because it discovers path only when needed and, therefore, eliminates periodic route information exchange as well as excessive routing overhead.

Second, which is the more important reason, table-driven routing poses a big challenge for designing load-sensitive routing protocols. In table-driven rout-

ing protocols, nodes exchange link cost information periodically and, therefore, the minimal cost path can be calculated at each node independently. To incorporate the idea of load-sensitive routing into a table-driven routing protocol, the link costs should reflect traffic load. However, with flows come and go, the traffic load at a node varies constantly. The link cost information obtained by periodical exchange may be outdated and, therefore, the calculated path may not be the current minimal cost one. For example, node i notifies its neighbors that it is idle. Thus, each node calculates that the path going through node i has the minimal cost. As flows are routed to node i one after another, node i's load increases gradually. Since node i could not notify other nodes that it is about to be overwhelmed before the next information exchange period, it would eventually be congested. During the next information exchange, after node i reports high traffic load, all the flows would be retreated from node i and be routed to another node that announce it is idle, leaving node i to be idle again. This technique results in route flapping and excessive routing overhead, which is a big challenge for designing load-sensitive routing protocols in wired networks [126] because most wired network routing protocols are table-driven. Fortunately, on-demand load-sensitive routing protocols could utilize up-to-date congestion information and effectively avoid this problem.

To facilitate our subsequent description, we use DSR to describe the idea of incorporating load-sensitive routing into existing on-demand protocols. We choose DSR because it is a typical on-demand routing protocol and, moreover, incorporating load-sensitive routing into other on-demand routing protocols is essentially the same as incorporating the idea into DSR.

7.2.3 Two Approaches to Discover Non-Congested Paths

Most on-demand routing protocols, including DSR, employ a blind flooding technique. The blind flooding technique requires every node to forward a fresh RREQ

as quickly as possible and discard subsequent duplicates. Thus, the discovered path has the minimum delivery latency for the fresh RREQ. However, the blind flooding process does not guarantee that the discovered paths circumvent hot spots.

To discover non-congested paths during the blind flooding process, we propose two straightforward approaches. One is to delay the propagation of the RREQ at a node according to the channel utilization, denoted as DPR (Delaying Propagating RREQ). In this way, we eliminate the chances for the congested nodes to forward the RREQ quickly. Another approach is to tag the RREQ if it is forwarded by a congested node so that the source would choose the path carried by the RREQ that is not tagged, denoted as TGR (Tagged RREQ).

DPR

DPR is almost identical to the basic blind flooding technique, except that, during each route discovery process, a node rebroadcasts the first arriving RREQ with a delay τ. τ is proportional to the node's channel utilization

$$\tau = chUtil * \epsilon \quad (seconds) \ , \tag{7.2}$$

where ϵ is a constant. The higher the channel utilization, the larger delay the RREQ experiences at the node. Thus, the first RREQ arriving at the destination carries the path that has the minimal accumulated load. Since the congested nodes would delay the RREQ propagation substantially, DPR effectively excludes them from the discovered paths.

Notice that (7.2) is a linear function. Other function forms, for example quadratic, are also reasonable potential choices. We leave this optimization issue as a future research topic.

TGR

In TGR, we choose a channel utilization threshold, $chUtil_{th}$, for the whole network. If a node's channel utilization is above $chUtil_{th}$, it is considered to be congested. Before a congested node forwards a RREQ, it should tag the packet so that the downstream nodes can determine whether the RREQ has traversed congested areas.

As in the blind flooding process, TGR requires a node to rebroadcast the fresh RREQ as quickly as possible so that paths between sources and destinations can be established quickly. Meanwhile, TGR allows a node to rebroadcast an un-tagged duplicate RREQ, if the previous RREQs it received are all tagged. Thus, the node broadcasts an RREQ at most twice during each route discovery process.

In this way, the un-tagged RREQ would be eventually forwarded to the destination and the source node could obtain a non-congested path if it exists.

A more sophisticated TGR scheme would be such that the RREQ counts the number of congested nodes it has traversed during its propagation. The source node then could choose the route that contains the least number of congested nodes. We leave exploration of this idea as potential future work.

7.3 Details of the Load-Sensitive Routing Protocols

Based on above considerations, we present the details of the two load-sensitive routing protocols, DPR and TGR, both of which are based on DSR.

7.3.1 DPR Protocol

DPR is essentially the same as DSR, except that, when a node receives a fresh RREQ, it rebroadcasts the RREQ with a delay τ calculated by (7.2).

Here, the key issue is to choose an appropriate value for ϵ in (7.2). If ϵ is too small, we could not effectively prevent the congested nodes from rebroadcasting

the RREQ quickly. If ϵ is too large, it would bring about substantial latency for path establishment and lead to lower network throughput as well as higher packet delivery latency. In our implementation, we set ϵ to 0.01 seconds, because, in the DSR implementation, each node rebroadcasts the RREQ with a delaying jitter less than 0.01 seconds to avoid collisions. Therefore, setting ϵ to 0.01 seconds should not impact the path establishment latency substantially and it is sufficient to prevent congested nodes from forwarding the RREQ faster than non-congested ones.

7.3.2 TGR Protocol

The key issue in TGR is to choose an appropriate channel utilization threshold, $chUtil_{th}$. A node is regarded as congested if its $chUtil$ is greater than $chUtil_{th}$. Either a large or a small $chUtil_{th}$ may prevent us from distinguishing congested nodes from non-congested nodes properly. In our implementation, we set $chUtil = 0.6$ because it gives us the best simulation results among other choices.

To incorporate the idea of TGR into DSR, we modify both the RREQ and RREP format by adding a block field, "BLK", into their packet header. When the source node broadcasts a fresh RREQ, it sets the "BLK" field in the RREQ to $false$. The "BLK" field is set to *true* if the RREQ is forwarded by a congested node.

Moreover, in DSR, each node maintains a Route Request Table (RRT). RRT is used to record the most recent route request for each source-destination pair and, therefore, a node can determine whether a received RREQ is fresh or a duplicate. The entry in the RRT contains three fields, *sid*, *did* and the most recent route request sequence number, seq_r. In TGR, the RRT has additional functionality. That is, during each route discover process, RRT should record whether the node has forwarded an un-tagged RREQ. Thus, the node could take different actions when receiving duplicate RREQs. For this purpose, we add an additional field, "fd_untag", into the RRT entry.

Whenever a source node, S, wants to discover paths to a destination, D, it broadcasts an RREQ with sequent number, s, and sets the "BLK" field to $false$. When a node receives the RREQ, it checks its $chUtil$. If $chUtil$ is greater than the threshold, $chUtil_{th}$, the node tags the RREQ by setting the "BLK" field to $true$. Thereafter, it retrieves the source node address, S, the destination address, D, and the sequence number, s, from the RREQ header. The node looks up its RRT table for the entry corresponding to triple $< S, D, s >$ and determines whether the RREQ is fresh or duplicate:

- If the RREQ is fresh, the node forwards it as quickly as possible no matter whether it is tagged or not. Meanwhile, the node updates the corresponding entry as follows. It sets $seq_r = s$. If the "BLK" in the RREQ is $false$, the node sets "fd_untag" to $true$, indicating that the node has forwarded an un-tagged RREQ during this discovery process; otherwise, it sets "fd_untag" to $false$.

- If the RREQ is a tagged duplicate, the node discards it without further actions. However, if the RREQ is an un-tagged duplicate, the node determines whether or not to forward the packet according to the "fd_untag" field in the entry. If "fd_untag" is $false$, which means the node has not forwarded an un-taged RREQ, the node forwards the RREQ as quickly as possible and set "fd_untag" to $true$. Otherwise, the node discards this duplicate request.

The RREQ is propagated until it arrives at the destination, which would then copy the route record and the "BLK" field into an RREP and sends the RREP back to the source along the path traversed by the RREQ. According to the "BLK" field in the RREP header, the source node knows whether or not the discovered path is congested.

If the source node receives multiple paths, it always chooses the one that is

not congested. If two paths have the same congestion status, the source node uses the number of hops to break the tie.

7.4 Reinvestigating the Blind Flooding Technique

During our simulations, we found that unmodified blind flooding has some ability to circumvent hot spots. This observation does not mismatch our initial impression that blind flooding tends to discover shortest-paths and motivates us to consider the technique more closely.

In the blind flooding process, every node forwards the fresh RREQ as quickly as possible and discards the subsequent duplicates. Thus, the discovered path is actually the one with the minimum delivery latency for the fresh RREQ. When network load is light, forwarding latency dominates the end-to-end packet delivery latency, and the path with the minimum delivery latency is the one with the least number of hops. However, when network load is high, queueing latency dominates the packet delivery latency, and the path with the minimal delivery latency may not be the one with the least number of hops. These observations imply two things:

- First, the blind flooding technique has some ability to bypass hot spots, because the RREQ, that is forwarded by congested nodes, may experience substantial latency due to intense channel contention and may not be able to arrive at destination in the first place. Particulary, when an area is significantly congested, it is very likely to be circumvented.
- Second, it is possible that the congested nodes may manage to forward the RREQ quickly because of possible instantaneous congestion relieve, resulting in congested paths being found first. This is especially likely, when the nodes are only moderately congested.

A second effect is also important. When the traffic load is high, transmissions

are more likely to fail due to intense channel contention, creating a route outage. This suggests that congested paths are more unstable than non-congested paths. If the blind flooding process discovers congested paths, route outages would occur frequently and source nodes would have to initiate route discoveries from time to time until they find non-congested paths that can be sustained for a relatively long duration. As a consequence, over time, the blind flooding technique tends to converge to paths in non-congested areas.

Obviously, once the blind flooding technique finds non-congested paths, it yields the same performance as load-sensitive routing protocols. If we amortize the performance degradation before the convergence over the entire communication period, the performance gap between blind flooding and load-sensitive routing protocols would decrease. When traffic flow life-time is short, the overall degradation may be substantial. But, when traffic flow life-time is long, the overall degradation would be trivial. In addition, node mobility has the same effect as flow life-time. As nodes move constantly, no paths can be sustained for a long time. The blind flooding process is not able to converge paths to non-congested areas and, therefore, the overall performance degradation may be substantial.

7.5 Simulation Results

We implemented the two load-sensitive routing schemes, DPR and TGR, in NS2 [5, 6] and evaluated their performance by comparing them with the unmodified blind flooding technique.

We use DSR to represent the blind flooding based routing protocols. Meanwhile, we disable two enhanced routing features provided by DSR. That is, an intermediate node is forbidden to reply with a RREP to the source and the destination is allowed to reply only to the first arriving RREQ.

We forbid the intermediate node from replying with RREPs, because we want

to prevent outdated congestion information from affecting current route discovery processes in DPR and TGR. For example, if an intermediate node knows a previously discovered "non-congested path", which is currently congested, to the destination and propagates it to the source, the source node may deliver data packets along this seemly non-congested but actually congested path. In fact, DSR also benefits from disabling this feature when hot spots exist [98].

We allow the destination to reply only to the first arriving RREQ, because, otherwise, DSR might easily obtain a sustainable non-congested path with very few route discovery efforts and the performance gap between DSR, DPR and TGR would be hard to observe. But, an exception is that, in TGR, the destination can reply with another RREP, if it receives an untagged RREQ and the previous RREQs were tagged.

Routing protocol performance is evaluated based on network throughput (or packet delivery ratio), packet delivery latency, and routing overhead. Packet delivery ratio is the number of data packets that arrive at the destinations divided by the number of data packets generated at the sources. It is equivalent to network throughput. Packet delivery latency is the time from when a data packet is generated at a source to the time when it is received at a destination. Routing overhead is the total number of bytes in the routing control packets that have been transmitted divided by the number of data packets received at the destinations. Therefore, it is a per data packet overhead.

We did extensive simulations with various traffic flows and topologies. However, the load-sensitive routing protocols do not improve network performance substantially in many cases. This was due to two reasons. First, as discussed earlier, the blind flooding technique is load-sensitive to a certain degree. Second, the performance of load-sensitive routing is affected by the actual traffic pattern. For example, if either the source or destination falls inside a hot spot, non-congested paths do not

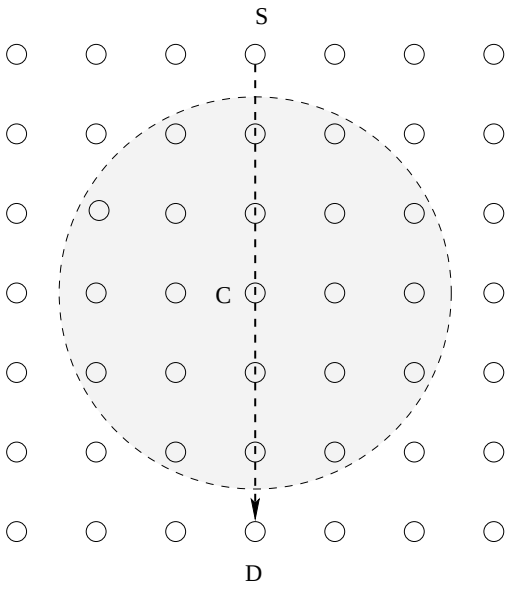


Figure 7.2: A grid topology with seven rows and seven columns of nodes.

exist and, thus, load-sensitive routing may not have impact on such scenario.

Nevertheless, as we pointed out in the previous section, the performance gap between blind flooding technique and load-sensitive routing may be impacted by flow life-time and node mobility. We perform simulations to confirm these arguments.

7.5.1 Flow Life-Time

We evaluate the impact of flow life-time on the performance of the load-sensitive routing and blind flooding technique using the grid topology in Figure 7.2.

Figure 7.2 shows a grid topology with seven rows and seven columns of nodes. Neighboring nodes are 200m apart. Node S is the source and node D is the destination. Node S sends data packets to node D at the rate of 4 packets/second. The packets' length is 512 bytes. Node C periodically sends out hello messages to simulate a congested area, which is illustrated by the dotted circle. The hello message interval at node C is t_h. The smaller t_h, the higher the congestion. The

transmit power at each node is P_{max} and thus the transmit range $250m$.

When presenting the simulation results, we use packet delivery ratio instead of network throughput. This is because the total number of data packets generated at the source S are relatively small. Any small jitter in data packet generation would result in slightly different network throughput. Because the performance for the three routing protocols is close, the slight jitter may skew our observations. Therefore, in this set of simulations, the packet delivery ratio is more illustrative than the network throughput.

Short-Life Flows

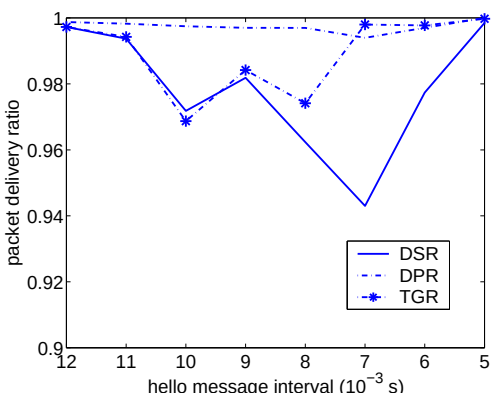
First, we evaluate the protocol performance for short life-time flows. We set the flow life-time between node S and D to 10 seconds and measure the protocol performance with varying t_h.

Figure 7.3(a) shows the packet delivery ratio. The x-axis is the hello message interval, t_h, at node C. We observe three regions:

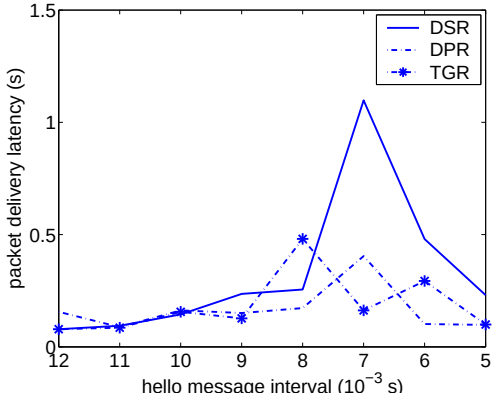
- When t_h is 0.012 seconds, DSR has the same packet delivery ratio as DPR and TGR, because area C is not congested and all the protocols route the traffic flow through area C.

- When t_h is 0.005 seconds, the three protocols demonstrate the same performance because area C is highly congested so that all protocols circumvent the hot spot and route the traffic flow through idle space.

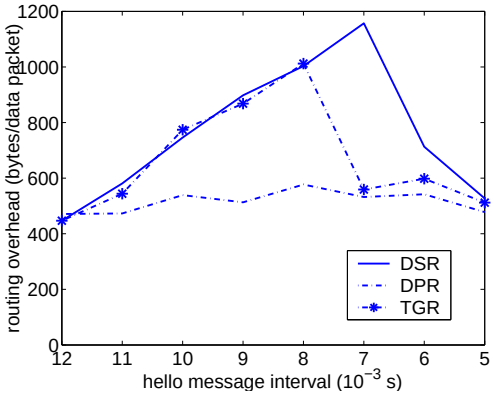
- When t_h is between 0.011 seconds and 0.006 seconds, in which case area C is moderately congested and DSR cannot circumvent the spot all the time, the three protocols are distinguished from each other. Specifically, TGR outperforms DSR when t_h is less than 0.008 seconds. This is because, when t_h is less than 0.008 seconds, the channel utilization within area C is greater than the



(a) packet delivery ratio



(b) packet delivery latency



(c) routing overhead per data packet

Figure 7.3: Performance of DSR, DPR and TGR for short-life traffic flows.

predefined threshold, $chUtil_{th}$, and TGR starts to circumvent this hot spot, which leads to higher packet delivery ratio. DPR always outperforms DSR because it is not triggered by a predefined threshold. Instead, DPR requires the nodes to forward RREQ with a delay proportional to their channel utilization. Therefore, DPR always routes the flow along the path that has the minimal traffic load, even when area C is slightly congested.

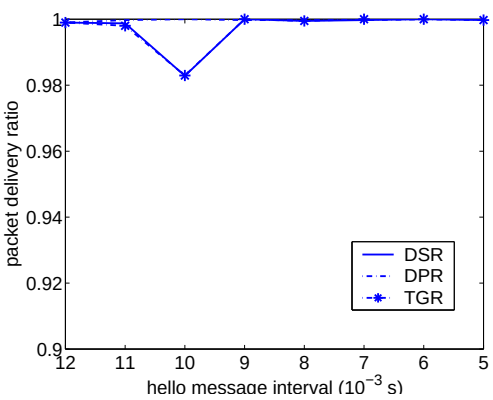
Figure 7.3(b) shows the packet delivery latency. For the same reason stated above, the three protocols have the same performance when t_h is either 0.012 seconds or 0.005 seconds. TGR outperforms DSR when t_h is less than 0.008 seconds, because it starts to circumvent the hot spot and, therefore, decreases the queuing delay at the forwarding nodes. DPR always outperforms DSR, because it circumvents area C even when area C is slightly congested.

Figure 7.3(c) shows the routing overhead. The y-axis is the per data packet routing overhead in bytes. Since non-congested paths are much more stable than congested paths, the load-sensitive routing protocols initiate fewer number of route discoveries and generate less routing overhead than DSR. When t_h is less than 0.008 seconds, TGR is triggered by the predefined threshold and starts to circumvent area C, resulting in less frequent route outages and lower routing overhead than DSR. Because DPR always routes the traffic flow through idle space, its routing overhead is lower than DSR in all cases.

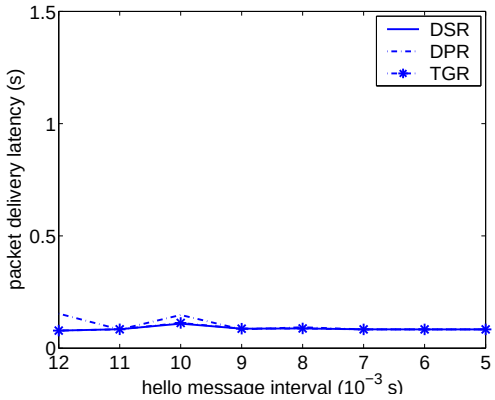
Long-Life Flows

To observe the protocol performance for long-life flows, we set the flow life-time between node S and node D to 1000 seconds and measure the network performance with varying t_h. Figure 7.4 shows the results. To compare with the results in Figure 7.3, we use the same scales on the y-axis.

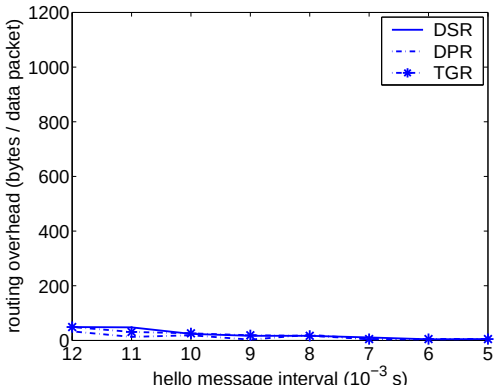
Figure 7.4(a) shows the packet delivery ratio. We observe that DSR has the



(a) packet delivery ratio



(b) packet delivery latency



(c) routing overhead per data packet

Figure 7.4: Performance of DSR, DPR and TGR for long-life traffic flows.

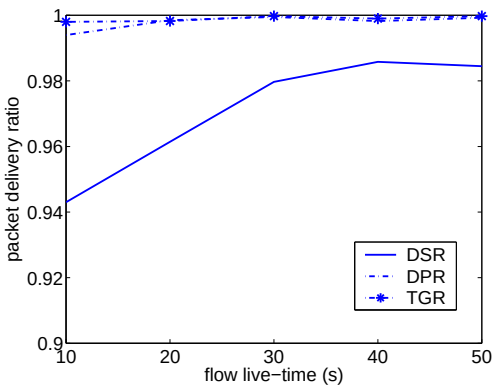
same performance with TGR and DPR, even when area C is moderately congested. This is because, if DSR discovers a congested path, it is likely to be broken immediately and the source node S would have to initiate another route discovery. This process continues until node S finds the non-congested path that would be sustained till the end of the communication session. Because the duration of the communication session is relatively long, the degraded performance at the beginning is amortized over the entire communication period. Therefore, the overall degradation is neglectable and we do not observe performance difference between DSR and the load-sensitive routing protocols.

For the same reason, the load-sensitive routing protocols have no substantial improvement on packet delivery latency and routing overhead, as shown in Figure 7.4(b) and (c).

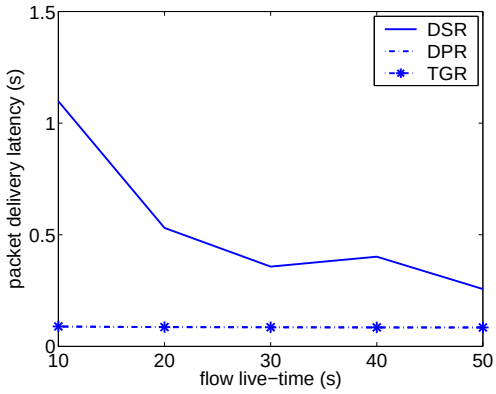
Varying Flow Life-Time

Finally, we present the protocol performance with varying flow life-time. To do this, we fix t_h at 0.007 seconds so that area C is moderately congested. We vary flow life-time from 10 to 50 seconds and measure the protocols performance. Because the traffic flow profile in multihop wireless networks is unclear, we resort to that in the internet. Claffy *et al.* [127] indicated that the majority of internet traffic flows are of tens seconds. For this reason, we choose a flow life-time in the range of 10 seconds to 50 seconds.

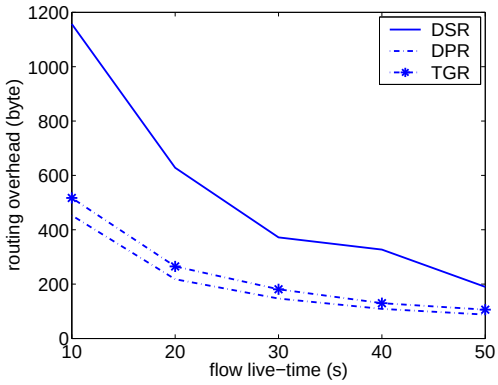
Figure 7.5(a) shows the packet delivery ratio. The x-axis is the flow life-time in seconds. We observe that both TGR and DPR outperform DSR. However, as the flow life-time increases, DSR gradually achieves the performance of the two load-sensitive routing protocols. This is because, DSR would eventually converge to paths in idle space. It underperforms the TGR and DPR only when it delivers data packets along congested paths. If DSR discovers non-congested paths, it will yield



(a) packet delivery ratio



(b) packet delivery latency



(c) routing overhead per data packet

Figure 7.5: Performance of DSR, DPR and TGR with varying flow life-time.

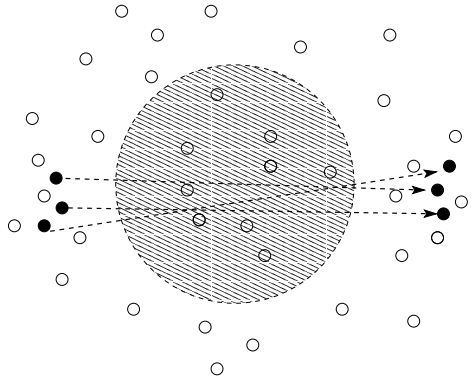


Figure 7.6: 90 mobile nodes are randomly put in an 1500m by 1000m area.

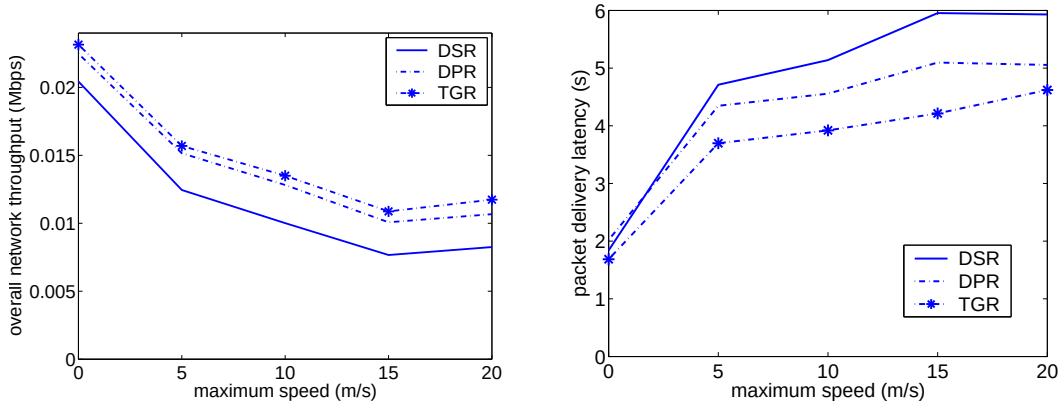
the same packet delivery ratio as the load-sensitive routing protocols. Thus, the initial performance degradation will be amortized over the entire communication period. When flow life-time increases, the overall degradation decreases.

Figure 7.5(b) and (c) show the packet delivery latency and routing overhead respectively. For the same reason stated above, as flow life-time increases, the performance of DSR gradually achieves the performance of the two load-sensitive routing protocols.

7.5.2 Node Mobility

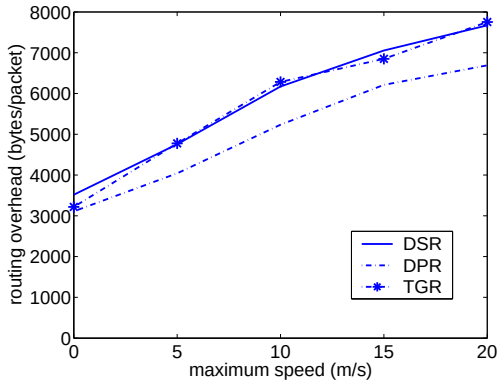
We evaluate the impact of node mobility on the protocol performance using the topology in Figure 7.6.

Figure 7.6 shows a topology, where 90 nodes are randomly put in an 1500m by 1000m area. The nodes move according to the random waypoint mobility model [6] at the speed not faster than v_{max}. A stationary node in the middle of the network broadcasts hello messages every 0.007 seconds so as to imitate a moderately congested area, as shown by the dotted circle. There are three traffic flows. The source nodes are stationary and they are on the left side of the congested area. The destinations are also stationary and they are on the right side of the congested area.



(a) network throughput

(b) packet delivery latency



(c) routing overhead per data packet

Figure 7.7: Performance of DSR, DPR and TGR with varying node mobility.

We intentionally arrange this traffic pattern to prevent either the sources or the destinations from falling into the congested area, in which case non-congested paths do not exist. Each source sends data packets at the rate of 4 packets/second and all packets are of length 512 bytes. The life-time of each flow is 300 seconds, which is the duration of the simulation. In doing this, we exclude the impact of short-lived flows and focus on the impact of node mobility. We measure the network performance with varying v_{max}.

Figure 7.7(a) shows the overall network throughput. The x-axis is v_{max}

in m/s and the y-axis is throughput in Mbps. Both TGR and DPR outperform DSR, because TGR and DPR attempt to circumvent the hot spot and deliver data packets along idle space. The throughput of TGR is slightly higher than that of DPR, because DPR delays the propagation of the RREQ at each node according to the node's channel utilization. Therefore, DPR has a higher path establishment latency, which results in slightly lower network throughput. Moreover, as v_{max} increases, the performance gap between DSR and the load-sensitive routing protocols increases slightly. This can be explained as follows. In the low mobility cases, the non-congested paths are sustainable. Once DSR finds the non-congested paths by chance, it yields network throughput as high as the load-sensitive routing protocols for a relatively long duration. However, in high mobility cases, no paths can be sustained for a long time due to the nodes constantly moving. Even if DSR finds a non-congested path, it may be broken soon and, during the next route discovery processes, DSR may go back to the congested paths again, resulting in more performance degradation than the load-sensitive routing protocols.

Figure 7.7(b) shows the packet delivery latency. Both TGR and DPR outperform DSR, because they route traffic flows through idle space and, therefore, decrease the queuing delay at the forwarding nodes. DPR has higher delivery latency than TGR, because DPR establishes the paths with higher latency. Moreover, for the same reason stated above, the performance gap between the load-sensitive routing protocols and DSR increases as v_{max} increases.

Figure 7.7(c) shows the routing overhead per data packet. DPR outperforms DSR for two reasons. First, DPR delivered a greater number of data packet to the destinations and, therefore, decreases the per data packet routing overhead. Second, because of discovering congested paths, DSR has to cope with frequent route outages and, therefore, generates a greater number of route discoveries and higher routing overhead. TGR generates about the same amount of routing overhead as DSR. This

is caused by the combined effect of several factors. Except for the two presented above, TGR allows duplicate RREQs to be forwarded, which leads to a greater number of routing control messages in each route discovery process and, therefore, higher routing overhead.

Chapter 8

Future Work

As future work, several issues are worth investigation. An obvious research topic is protocol optimization. For protocols like EOTP, MHR and TGR, we did not attempt to identify the best possible functional forms of the equations and the best values of the factors that they employ. Although the values that we use are those that yield the best network performance in our simulations, it might be worthwhile to identify the optimal choices using formal analysis. However, since the optimal choices may depend on and vary with specific network scenario, it would be best to investigate the optimization issue after we have a clear understanding on the key characteristics of multihop wireless networks.

In addition, there are two broad categories of potential research topics. The first category is addressing the limitations of this dissertation such as rate adaptation, fast fading, etc. The second category are extensions of the studies in this dissertation such as combining mini-hop routing and load-sensitive routing. In the following sections, we present some insights about these issues and suggest potential research topics.

8.1 Rate Adaptation

From the spatial usage point of view, power control improves spatial utilization by reducing the reserved space of each individual transmission and thus allowing more concurrent transmissions, while rate adaptation improves spatial utilization by reducing the duration of each space reservation and thus allowing more transmissions to be completed in a time period. The two approaches pursue efficient spatial utilization along two different directions: space efficiency and time efficiency.

In the previous chapters, we discussed the impact of rate adaptation on our work and proposed several potential research topics. For example, designing a scheme that switches between power control and rate adaptation according to network traffic load.

8.2 Fast Fading

In this dissertation, we did not consider the effect of fast fading. Fast fading causes receive signal power to vary around a level determined by the large propagation model [8]. It impacts the ideal three-range model introduced in Chapter 2 and 4 such that the three ranges, transmit range, carrier sense range and interference range, vary constantly around the standard circles defined by the large propagation model.

8.2.1 Fast fading and OTP

OTP adjusts transmit power, P_{otp}, such that the carrier sense range is just large enough to cover the interference range. If we were to consider fast fading, the carrier sense range and the interference range may deviate from the standard circles defined by the ideal three-range model. P_{otp} may not be able to create the carrier sense range that exactly cover the interference range. Conservatively, we can add

a power margin to P_{otp} so as to prevent the interference range from exceeding the carrier sense range due to fast fading. But, since our study shows that OTP is overly conservative already, we might simply leave P_{otp} alone.

Moreover, OTP adjusts the transmit power of each individual data packet according to the immediately preceding RTS/CTS exchange and the RTS/CTS exchange is always performed at the maximal power level. When channel coherence time is in the magnitude of several data packet transmissions, OTP reacts quickly to channel variations. Thus, we would expect fast fading not to have much impact on OTP.

8.2.2 Fast Fading and EOTP

EOTP requires a sender to observe the worst SINR at a receiver for a relatively long interval, T_o, and then adjust transmit power, P_{eotp}, accordingly. Thus, the sender captures both the typical bad channel conditions caused by fast fading and the typical high co-channel interference.

When channel coherence time is long such that the channel varies more slowly than the interference envelope does, EOTP would react quickly enough to the channel condition variation. In this case, fast fading has little affect on EOTP and the performance of EOTP would be similar to that when we only consider large propagation model.

When channel coherence time is short, the sender may tune transmit power to a high level to cope with frequent bad channel condition, which causes the performance to be lower than when we only consider the large propagation model. To improve the performance, the sender could use power control to cope with the long-time scale interference envelope variation and use rate adaptation to cope with short-time scale channel condition variation. For example, during an observation period T_o, the sender transmits at a certain power level P_{eopt}, it leverages RTS/CTS

exchange to monitor the channel condition and varies transmission rate accordingly. Now, the question is "why does not the sender simply change P_{eotp} faster to cope with the channel condition variation?" This is because P_{eotp} should be varied to reflect the typical low SINR at the receiver. Since transmit power determines potential co-channel interference, every time when the sender changes the transmit power, it needs a certain duration to observe the impact of this change on the interference envelope at the receiver. Therefore, the sender can determine subsequent transmit power appropriately. If P_{eotp} is varied too fast, the sender does not have enough time to estimate the interference envelope, which would introduce lot of uncertainties to reliable signal receptions.

Intuitively, channel coherence time may affect T_o, which should be determined such that the sender is able to capture the typical bad channel condition and the typical high interference level at the receiver. Therefore, T_o is related to both channel coherence time and traffic pattern around.

8.2.3 Fast Fading and MHR

MHR uses spatial usage as a route selection metric. The spatial usage of a path is the accumulation of d^2 at each hop, where d is sender-receiver distance. Since we derive d according to the transmit and receive power levels, fast fading introduces errors. The error margin of each hop-distance is aggregated in the path's spatial usage metric, which may affect route selection. Because each hop-distance has chances to be overestimated, underestimated or maybe correctly estimated, the aggregated error margin may relate to the number of hops along the path. Understanding this relationship would help us choose route appropriately.

It needs to be mentioned that d could be estimated more accurately using sophisticated physical layer schemes or techniques such as GPS and triangulation.

8.3 Analyzing Spatial Usage with Other MAC Models

The analysis in Chapter 3 is based on the CSMA random access mechanism. But, CSMA is not an ideal way of scheduling transmissions. It tries to eliminate co-channel interference at the receiver by prohibiting nodes around transmitter from transmitting simultaneously. Therefore, some nodes that are close to the receiver but far away from the transmitter could still commence concurrent transmissions and corrupt the signal at the receiver. On the other hand, some nodes that are far away from the receiver but close to the transmitter have to defer their own transmissions, even though their concurrent transmissions would not corrupt the signal at the receiver.

An ideal MAC mechanism should be such that transmissions are scheduled on the basis of the interference level at the receivers. There are two models to describe the ideal MAC mechanism. The first one comes from [42] and can be described as follows:

Definition 6 *Let Tx denote the transmitter, Rx denote the desired receiver, and Tx_i denote a concurrent transmitter other than Tx. The signal is successfully received by Rx if:*

1. *The distance between Tx and Rx is less than R_{tr}, i.e.,*

$$|Tx - Rx| \leq R_{tr} \ . \tag{8.1}$$

2. *For every concurrent transmitters, the distance between Tx_i and Rx satisfies*

$$|Tx_i - Rx| \geq \Delta R_{tr} \ , \tag{8.2}$$

where R_{tr} is the transmit range, and Δ is a constant for any Tx_i.

The first condition specifies that the distance between the transmitter and receiver should not greater than R_{tr}. The second condition specifies that there is a constant interference range (IR) around the receiver within which no other transmitters can transmit simultaneously. In this model, the IR corresponds to the worst case where the transmitter and receiver are R_{tr} distance apart. The size of the IR does not vary with the T-R distance. If considering only one interference source, $\Delta = \sqrt[\alpha]{SINR_0}$ according to (2.7), where α is the path loss exponent, and $SINR_0$ is the capture effect threshold. If considering aggregate co-channel interference, $\Delta > \sqrt[\alpha]{SINR_0}$. To choose an appropriate value for Δ is part of our future work.

In reality, the interference tolerance level at the receiver actually varies with the T-R distance. The closer the transmitter and the receiver, the higher the interference level the receiver can tolerate. Based on the first model, we propose a more realistic one:

Definition 7 *Let Tx denote the transmitter, Rx denote the desired receiver, and Tx_i denote the concurrent transmitter other than Tx. The signal is successfully received by Rx if:*

1. *The distance between Tx and Rx is less than R_{tr}, i.e.,*

$$|Tx - Rx| \leq R_{tr} \ . \tag{8.3}$$

2. *For every concurrent transmitters, the distance between Tx_i and Rx satisfies*

$$|Tx_i - Rx| \geq \Delta |Tx - Rx| \ , \tag{8.4}$$

where R_{tr} is the transmit range, and Δ is a constant.

Here, the size of the IR varies with the T-R distance. Δ is $\sqrt[\alpha]{SINR_0}$ if there is only one interference source. It is greater than $\sqrt[\alpha]{SINR_0}$ if there is aggregate co-channel

interference. In this model, the size of the IR varies with the T-R distance, which results in the difficulties in identifying the spatial usage of each single transmission.

8.4 Improving TGR

The TGR protocol we implemented in Chapter 7 employs a mechanism that simply indicates whether a path is congested or not. A more sophisticated mechanism would be such that the RREQ counts the number of congested nodes as it is propagated. Thus, a source node could choose the route that contains the least number of congested nodes. This sophisticated TGR is similar to DPR in the way that both accumulate the congestion status at each hop and then choose the least congested path. Because TGR does not delay RREQ propagation and thus establishes paths more quickly than DPR, we expect the sophisticated TGR to outperform DPR but the increased performance may not be substantial.

8.5 Combining Mini-Hop and Load-Sensitive Routing

Originally, we planned to combine our two routing strategies, mini-hop routing and load-sensitive routing, and to investigate their synergistic performance.

However, Chapter 7 shows that the blind flooding technique, which is the basis of most on-demand wireless routing protocols, is load-sensitive in many cases. Although we enhanced the technique to enforce load-sensitivity and developed two protocols, DPR and TGR, we did not observe substantial improvement in network performance. Specifically, DPR and TGR outperform basic blind flooding only when traffic flow lifetime is relatively short or node mobility is relatively high.

Because the mini-hop routing protocol, MHR, is based on blind flooding, it should also have some level of load-sensitivity. It is likely that the combination of mini-hop routing and load-sensitive routing may not be able to outperform the

combination of mini-hop routing and blind flooding significantly. If we combine mini-hop and load-sensitive routing, some issues need to be considered. For example, the priority of the metrics in route selection, spatial usage or congestion status?

Chapter 9

Contributions and Conclusions

In this dissertation, we analyzed the impact of transmit power on the spatial usage of multihop transmissions and potential network throughput. Motivated by the analysis, we developed MAC and network layer protocols that leverage power control to improve network spatial utilization and performance. More importantly, we learned some lessons that are fundamental to the power control technique and useful in future protocol design and network deployment.

9.1 Basic Contributions

In Chapter 3, we proposed a spatial usage metric, Ω, and bounded it as a function of transmit power. Because Ω indicates potential network throughput, the impact of transmit power on potential network throughput was disclosed. Our study shows that the potential network throughput is bounded by two linear functions of $\frac{1}{R_{cs}}$ in 2D space and bounded by two constants in 1D space, where R_{cs} is the radius of the carrier sense range of the power level at which packets are transmitted. It suggests that the potential network throughput increases as transmit power decreases in 2D space but does not change with respect to transmit power in 1D space. After

applying two well-known approaches for eliminating the border effect, our simulation results match the theoretical analysis.

In Chapter 4, taking advantage of the insights gained by the spatial usage analysis, we developed a MAC layer power control scheme, OTP, to balance the spatial usage and co-channel interference of each individual transmission. Unlike most existing power control schemes that employ a dedicated control channel to protect transmissions on the data channel, we base our design on a single communication channel. To guarantee reliable signal reception, OTP adjusts transmit power based on the worst case interference. Our simulations demonstrate a moderate improvement over the 802.11 DCF MAC.

In Chapter 5, motivated by the observation that the worst case interference does not occur all the time, we enhanced OTP and developed EOTP which adjusts transmit power more aggressively. EOTP requires a sender to observe the actual interference level at a receiver for a relatively long period and then adjust subsequent transmit power accordingly. Thus, EOTP trades off a possible occasional collision for lower power and better spatial usage. The simulations show that EOTP outperforms both OTP and 802.11 significantly.

In Chapter 6, to take more advantage of the MAC layer power control, we investigated mini-hop routing, which discovers routes consisting of short hops, and developed a mini-hop routing protocol, MHR. Most existing mini-hop routing protocols use a critical power, P_c, to control route discovery processes so as to exclude long hops. They maintain P_c at the cost of substantial routing overhead. To reduce the overhead, we identify P_c according to previously discovered paths. To do this, we employ link-cost routing in each individual route discovery process so as to discover the minimum cost path with respect to spatial usage and thus reveal the minimum connectivity between a source and a destination. P_c is the power level that can cover the longest hop in the minimum cost path. It will be used to control the next

route discovery process for the same source-destination pair. Therefore, MHR is a combination of link-cost routing and the power control method. Our simulations show that MHR combined with EOTP improves network performance substantially while still keeping routing overhead low.

Finally, in Chapter 7, to fully utilize the networks spatial resource, we investigated load-sensitive routing, which bypasses hot spots and routes traffic flows into idle space. We found that blind flooding is load-sensitive to a certain degree. To reinforce load-sensitivity, we enhanced the blind flooding technique and developed two routing protocols: DPR and TGR. In DPR, a node delays the RREQ propagation for a duration that is proportional to its congestion degree. Therefore, the RREQ forwarded by congested nodes is not likely to arrive at the destination in the first place. In TGR, the RREQ is tagged if it is forwarded by a congested node. Therefore, a source node can distinguish non-congested paths from congested ones. The simulations show that DPR and TGR outperforms the blind flooding technique only when traffic flow lifetime is relatively short or node mobility is relatively high.

9.2 High Level Lessons Learned

In wireless networks, space is limited resource. Transmit power level determines the amount of space each individual transmission consumes and thus the space contention degree. For this reason, power control impacts many aspects of network performance.

In addition to above basic contributions, we learned valuable lessons and understood many aspects of this technique, including the applicable network scenarios, its impact on packet delivery latency and routing overhead, the appropriate power adjustment, etc. These lessons can be used as guidance for developing other power control protocols or deploying wireless networks.

9.2.1 Network Scale

An important insight is that power control is unlikely to help in small scale networks. When the network scale is small, either transmit power cannot be reduced to a level that is low enough to decouple transmissions, or the space conserved by reducing transmit power falls mostly outside of the network area and could not be utilized by other transmissions.

In particular, when MAC layer power control is combined with mini-hop routing, their synergistic operation actually degrades small scale network performance. This is because mini-hop routing reduces hop distances at the cost of a greater number of hops. If the hop distances are not short enough, the MAC layer power control scheme brings about no benefit. Meanwhile, mini-hop routing requires more transmissions to deliver each data packet and, therefore, increases channel contention. Thus, the combined effect degrades network performance.

The success of power control depends on how the conserved space is utilized by other transmissions. It is not applicable for small scale networks. It would be desirable to design a MAC layer protocol that can adapt to network scale and disable/enable power control accordingly. Also, it would be desirable to design a sophisticated routing protocol that switches between shortest path routing and mini-hop routing according to specific network topology and traffic pattern.

9.2.2 1D Topologies vs. 2D Topologies

Our study of 1D topology and 2D topology reveals the spatial usage within a single traffic flow and between multiple competing traffic flows.

Chapter 3 shows that the maximum network capacity is proportional to $\frac{1}{R_{cs}}$ in 1D space and $\frac{1}{R_{cs}^2}$ in 2D space, where R_{cs} is the radius of the carrier sense range of the power level at which packets are transmitted. Thus, power control increases network capacity more aggressively in 2D space than in 1D space. Moreover, our

study shows that the combination of MAC layer power control and mini-hop routing improves 2D networks performance but has no impact on 1D networks, which suggests that the synergistic operation may not be able to improve the spatial utilization within a single traffic flow but can improve the spatial utilization between multiple competing traffic flows.

Therefore, it is desirable for MAC layer power control protocols and network layer routing protocols to adapt to specific network topologies.

9.2.3 Packet Delivery Latency

Although our initial intent was to use power control to improve network throughput, we found that this technique improves other aspects of network performance, such as end-to-end packet delivery latency and in some cases routing overhead.

Packet delivery latency consists of forwarding latency and queuing latency. Forwarding latency is determined by the number of hops a packet undergoes and queuing latency is determined by the channel contention at each forwarding node. When the load is light, queuing latency is at a minimum and forwarding latency dominates the packet delivery latency. However, when the load is high, queuing latency increases significantly and dominates the packet delivery latency. Since power control reduces channel contention when traffic load is high, it reduces queuing latency and thus packet delivery latency.

Moreover, mini-hop routing helps MAC layer power control reduce channel contention further and thus reduce queueing latency to an even lower level. Although forwarding packets along mini-hop routes increases forwarding latency due to the fact that a mini-hop route consists of a greater number of hops, the combined effect is that the end-to-end packet delivery latency is reduced further.

An extra benefit of reducing channel contention is that established routes are more stable, resulting in less frequent route discoveries and lower routing overhead.

We realized that power control improves many aspects of network performance because it reduces channel contention. Therefore, any technique that reduces channel contention may be able to improve network performance.

9.2.4 Appropriate Power Adjustment

We found that mini-hop routing does not work well with conservative MAC layer power control schemes. When we begun, we attempted to show that the combination of OTP and MHR improves network performance. But, we did not obtain the expected results. Investigation shows that OTP adjusts transmit power in an overly conservative manner so that the benefit of delivering packets along short hops is nullified by the cost of delivering packets along a great number of hops.

Realizing that the worst case interference assumed by OTP rarely occurs due to low probability of the worst interferer and a sender obtaining the channel simultaneously, we developed EOTP, which tunes transmit power based on the actual interference level at a receiver. Although transmitting data packets at a power level that is determined by previous interference information introduces some risks, the chances of a packet being corrupted is small as long as the sender observes the interference at the receiver for a sufficiently long interval. Our simulation shows that EOTP outperforms OTP and the synergistic operation of EOTP and MHR improves the performance of using EOTP alone.

Thus, when designing power control protocols, it is appropriate to trade off a possible occasional collision for lower power and better spatial usage.

9.2.5 Mobility Resistance

There is a common belief that mini-hop paths are more vulnerable to node mobility than shortest paths [128], because the frequency of route outage may increase as the number of hops increases. However, our investigation shows that, in fact, mini-hop

paths may be more resistant to node mobility.

Every network defines a maximum transmit power and thus a maximum hop distance. When the sender and the receiver move beyond the maximum distance, the hop is broken, resulting in outages for the routes that contain this hop.

Shortest-path routing favors long hops so that a path contains as few hops as possible. Therefore, each hop tends toward the maximum allowable distance. When nodes move, each hop is likely to be broken. However, mini-hop routing favors shorter hops. If the MAC layer power control scheme is adaptive to the sender-receiver distance, each short hop can be extended to the maximum distance and is, therefore, more resistant to breaking due to node mobility.

Although short hop-distance leads to fewer route outages, we should also realize that the frequency of route outage may increase as the number of hops increases in a path. Our simulations in Chapter 6 show that, as node mobility increases, mini-hop routing brings about lower routing overhead than shortest-path routing. Although it is needed to perform a formal investigation on this issue, our study suggests a good side of mini-hop routing in terms of reducing routing overhead.

Therefore, if we want to design a routing protocol that discovers routes with high mobility resistance, we may employ a strategy that chooses hops with short distance.

9.2.6 Controlling Transmit Power in the Route Discovery Processes

To discover mini-hop paths, some researchers proposed to control transmit power during the route discovery process [113, 114, 115]. In this approach, the key issue is to find the critical power, P_c, that is high enough to keep the network connected but low enough to exclude long hops. The initial intuition is that the lower P_c the better. But, our investigation shows that, when P_c is so low that it means the network is

barely connected, per data packet routing overhead increases tremendously.

To explain this, we need to understand route caching. As introduced in Chapter 2, many wireless routing protocols allow a source node to maintain multiple paths to the same destination. When one path is broken, the source node uses another without initiating a route discovery. This process continues until no path is available in the cache. In this way, the frequency of route discovery is decreased and routing overhead is reduced significantly.

However, as P_c reduces, network connectivity is weakened, resulting in fewer number of potential paths between a source and a destination. During each route discovery process, there are fewer paths that can be discovered and be cached at the source. Route discoveries may occur more frequently, resulting in higher routing overhead.

Therefore, when using the power control approach to discover mini-hop routes, the power level should be chosen carefully so as to avoid significant routing overhead. In particular, the power level that barely keep network connected may not be an appropriate choice.

9.2.7 Load-Sensitivity in the Blind Flooding Technique

Our investigation in Chapter 7 shows that blind flooding technique is load-sensitive to a certain degree. This is caused by two reasons.

The first reason is the nature of blind flooding. In the basic blind flooding route discovery process, a node is required to forward the first arriving RREQ as quickly as possible and discard the subsequent duplicates. Therefore, the discovered path has the minimum delivery latency for the RREQ. If the RREQ is forwarded by a congested node, it would experience significant queueing latency and may not be able to arrive at the destination first. However, it is still possible that a congested node may manage to forward the RREQ quickly, which leads to a congested path

being discovered. This is more likely when the node is moderately congested.

The second reason is the nature of wireless communications and 802.11 based MAC protocols. When channel contention is intense, transmissions are likely to fail due to corruptions and collisions. In 802.11, when the number of consecutive transmission failure exceeds a predefined threshold, the sender and receiver is considered unreachable and, therefore, a route error message will be reported to the source node, resulting in a route outage. Therefore, congested paths are much more likely to be broken than non-congested path.

The above two facts cause the blind flooding technique to converge traffic flows into idle space eventually if the traffic flow lifetime is long. This suggests that the blind flooding technique may be good enough for load-sensitive routing in network scenarios that traffic flow lifetime is long and node mobility is low.

9.3 Conclusions

In this dissertation, we designed MAC and network layer power control protocols and showed that they improve wireless network spatial reuse, resulting in overall network performance.

Appendix A

Proof of Two Lemmas about Shortest-Path Routing

Now, we prove the two lemmas presented in Chapter 2. They are used in Chapter 3 when we analyze the impact of transmit power on spatial usage.

Lemma 1 *Let R_{tr} denote the transmit range of the power level at which a shortest path is discovered. The average hop-distance of two consecutive hops along the shortest path, $\overline{d}_2$, is bounded by*

$$0.5R_{tr} < \overline{d}_2 < R_{tr} \ . \tag{A.1}$$

Proof: Consider Figure A.1. Node i, $i+1$ and $i+2$ are three consecutive forwarding nodes along a shortest path. Let $d_{i,i+1}$, $d_{i+1,i+2}$ and $d_{i,i+2}$ denote the distance from i to $i+1$, $i+1$ to $i+2$ and i to $i+2$ respectively. The circles with radius R_{tr} are the transmit ranges of i and $i+2$ respectively. To guarantee reliable transmission, we have

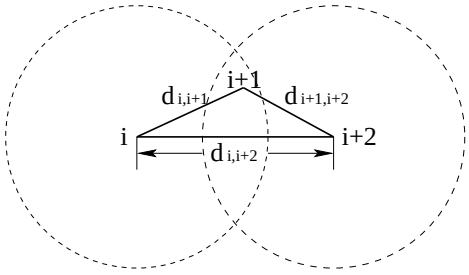


Figure A.1: Hop-distances in the shortest path routing.

$$\begin{cases} 0 < d_{i,i+1} \leq R_{tr} \\ 0 < d_{i+1,i+2} \leq R_{tr} \end{cases} . \tag{A.2}$$

Moreover, we know

$$R_{tr} < d_{i,i+2} \ , \tag{A.3}$$

because, otherwise, the shortest path routing would short-cut node $i+1$ and send packets from i to $i+2$ directly. According to the triangle inequality, $d_{i,i+2} \leq d_{i,i+1} + d_{i+1,i+2}$, we have

$$R_{tr} < d_{i,i+1} + d_{i+1,i+2} \leq 2R_{tr} \ . \tag{A.4}$$

Thus, the average hop-distance $\overline{d}_2$ is bounded by

$$0.5R_{tr} < \overline{d}_2 \leq R_{tr} \ . \tag{A.5}$$

Lemma 2 *Let L denote the source-destination distance, and R_{tr} denote the transmit range of the power level at which a shortest path is discovered. The number of hops on the shortest path is a function of R_{tr}, denoted by $H(R_{tr})$. We have the following three inequalities*

$$H(R'_{tr}) \leq H(R_{tr}), \qquad (R'_{tr} > R_{tr}) \ , \tag{A.6}$$

$$H(2R_{tr}) \leq \frac{1}{2}H(R_{tr}), \qquad (2R_{tr} \leq L) \ , \tag{A.7}$$

and

$$H(nR_{tr}) \leq \frac{1}{n}H(R_{tr}), \qquad (n = 1, 2, 3, ..., \ nR_{tr} < L) \ . \tag{A.8}$$

Proof: The first inequality, we prove by contradiction. Let P_t and P'_t be the power level corresponding to R_{tr} and R'_{tr} respectively. If $H(R'_{tr}) > H(R_{tr})$, the path discovered at P'_t is not the shortest because the path discovered at P_t is also reachable for P'_t. This violate the condition of shortest path routing. Thus, the first equation is proven.

For the second inequality, let us consider the two consecutive hops in Figure A.1. If the transmit range increases to $2R_{tr}$, i can reach $i + 2$ directly and shortcut node $i+1$. It is even possible that i may shortcut more hops, depending on the positions of the subsequent forwarding nodes. Thus, when the transmit range doubles, we can construct a new path by replacing two consecutive hops with a single hop. The number of hops of this constructed path is $\frac{1}{2}H(R_{tr})$. Since the shortest path contains the minimal number of hops, $H(2R_{tr}) \leq \frac{1}{2}H(R_{tr})$.

For the same reason, if the transmit range increases to nR_{tr}, a new path can be constructed such that a single hop replaces at least n original hops. The number of hops of this constructed path is $\frac{1}{n}H(R_{tr})$. Because of the shortest path routing, $H(nR_{tr}) \leq \frac{1}{n}H(R_{tr})$.

Note that nR_{tr} should be less than the S-D distance L. Otherwise, $H(nR_{tr}) = 1$ and the above equations may not be valid.

Appendix B

Hop-Distance Distribution in Shortest-Path Routing

We want to obtain some insights into the average hop-distance $\overline{d}$ and hop-distance distribution of shortest-path routing by using Monte Carlo simulations. Understanding the hop-distance distribution is important for spatial usage analysis. In Chapter 4, we use it to analyze the performance of OTP. In addition, it is useful in localization and distance estimations [116].

We simulate both 1D and 2D networks, and all simulations are performed in ns2 [5, 6]. We randomly put n nodes into a fixed size network with area S. For 1D space, $S = 1000m$. For 2D space, $S = 1000m \times 1000m$. Further, 20 Source-Destination (S-D) pairs are randomly chosen from these nodes. The routes for these S-D pairs are discovered by the shortest-path routing protocol, SRP, at power level P_t whose corresponding transmit range is R_{tr}. We measure each hop distance d along the shortest pathes. The possible values of SRP, n and R_{tr} are:

- $SRP \in [DSR, AODV, DSDV]$
- $n \in [80\ nodes, 120\ nodes, 160\ nodes]$ for 1D space, and, $n \in [120\ nodes, 160\ nodes, 200\ nodes]$

for 2D space.

- $R_{tr} \in [250m, 225m, 200m, 175m, 150m, 125m, 100m]$

Thus, each simulation setup is a combination of $<SRP,\ n,\ R_{tr}>$, and is repeated 30 times. It yields a set of measured hop-distances from which the hop-distance distribution and $\overline{d}$ are obtained.

B.1 Experiment in 1D Space

Figure B.1 shows the variations of $\overline{d}$ with respect to (w.r.t) R_{tr} in 1D space. The x-axis is R_{tr}, and the y-axis is $\overline{d}$. The discrete points are the experimental results, and the lines are the curves of

$$\overline{d} = 0.75R_{tr}\ . \tag{B.1}$$

We observe that, for DSR, AODV and DSDV, the measured $\overline{d}$ is very close to $0.75R_{tr}$, and it is affected very little by node density.

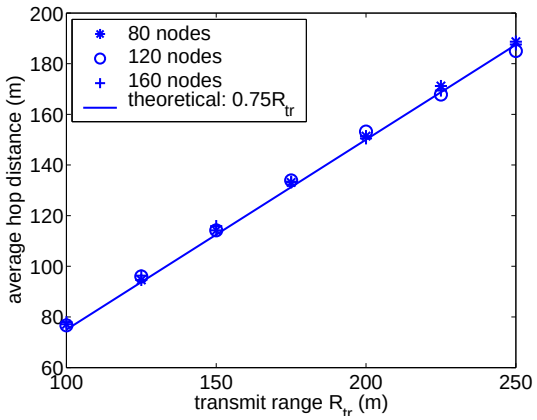
Figure B.2 shows the cumulative distribution of the hop-distance for DSDV. The x-axis is the hop-distance, d, and the y-axis is the cumulative probability. The discrete points are the experimental results, and the lines are curves of

$$cdf(d) = e^{-\frac{4}{R_{tr}}(R_{tr}-d)} \qquad (0 \leq d \leq R_{tr})\ . \tag{B.2}$$

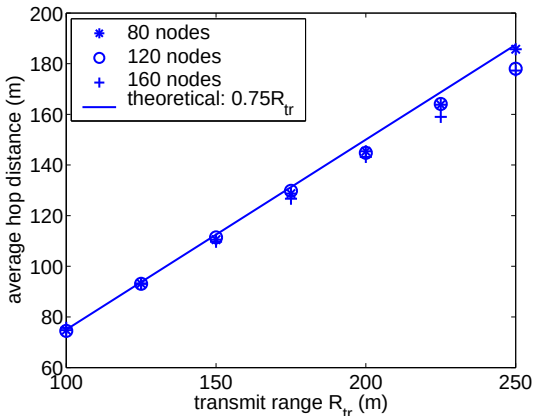
This equation is an exponential distribution with $\overline{(R_{tr}-d)} = \frac{R_{tr}}{4}$, or equivalently $\overline{d} = 0.75R_{tr}$. We observe that the hop-distribution is very close to (B.2).

B.2 Experiment in 2D Space

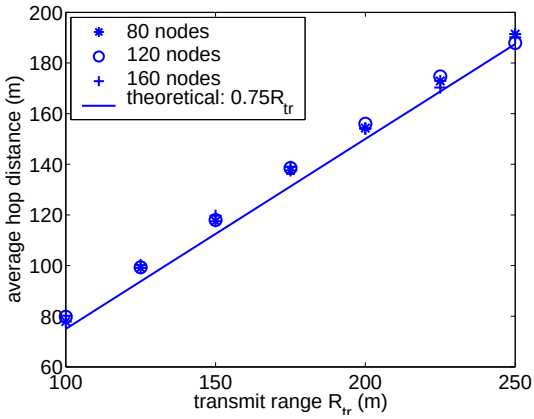
Figure B.3 shows the variations of $\overline{d}$ w.r.t R_{tr} in 2D space. The x-axis is R_{tr}, and the y-axis is $\overline{d}$. The discrete points are the experimental results, and the lines are



(a) DSR

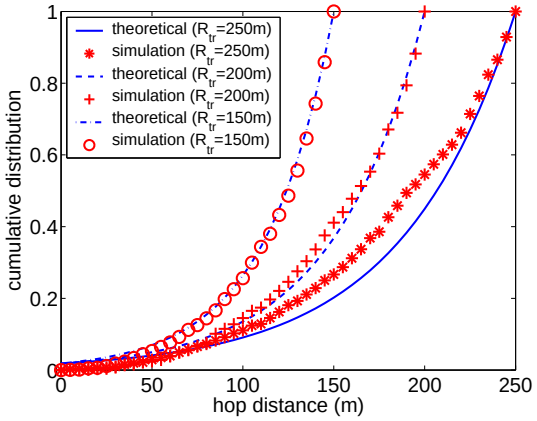


(b) AODV

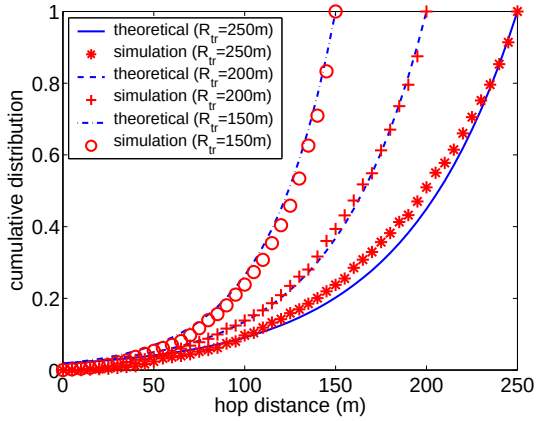


(c) DSDV

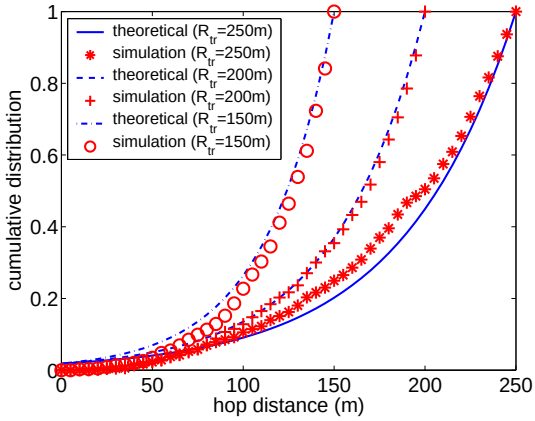
Figure B.1: Average hop distance vs. transmit range in 1D space.



(a) DSDV, n=80



(b) DSDV, n=120



(c) DSDV, n=160

Figure B.2: Hop-distance distribution in 1D space.

the curves of

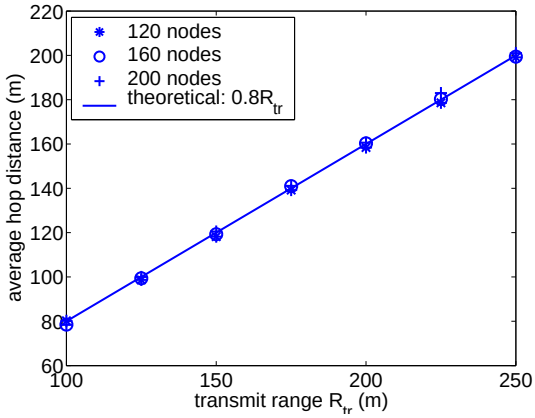
$$\overline{d} = 0.8R_{tr} \ . \tag{B.3}$$

We observe that, for DSR, AODV and DSDV, the measured $\overline{d}$ is very close to $0.8R_{tr}$, and it is affected very little by node density.

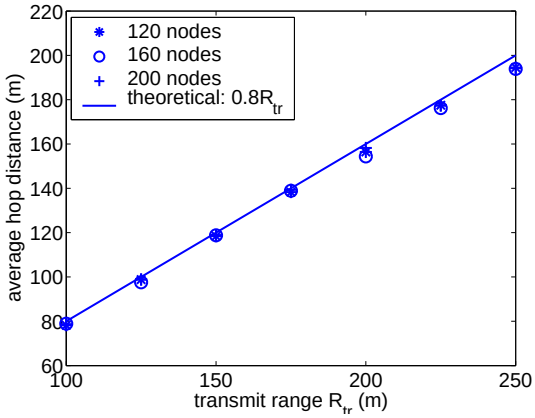
Figure B.4 shows the cumulative distribution of the hop-distance for DSDV. The x-axis is d, and the y-axis is the cumulative probability. The discrete points are the experimental results, and the lines are curves of

$$cdf(d) = e^{-\frac{5}{R_{tr}}(R_{tr}-d)} \qquad (0 \leq d \leq R_{tr}) \ . \tag{B.4}$$

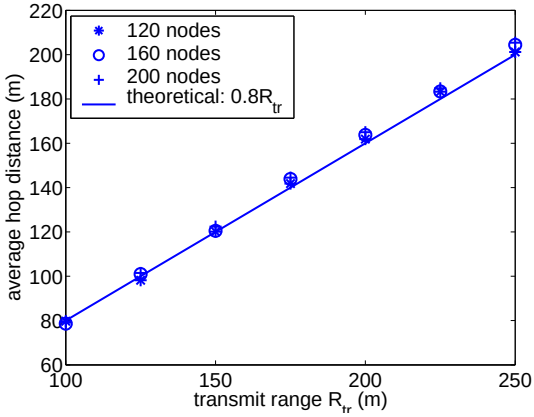
This is a exponential distribution with $\overline{(R_{tr} - d)} = \frac{R_{tr}}{5}$, or equivalently $\overline{d} = 0.8R_{tr}$. We observe that the hop-distance distribution is very close to the function from (B.4).



(a) DSR

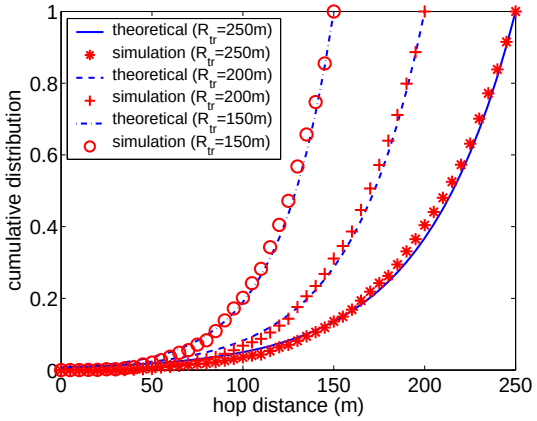


(b) AODV

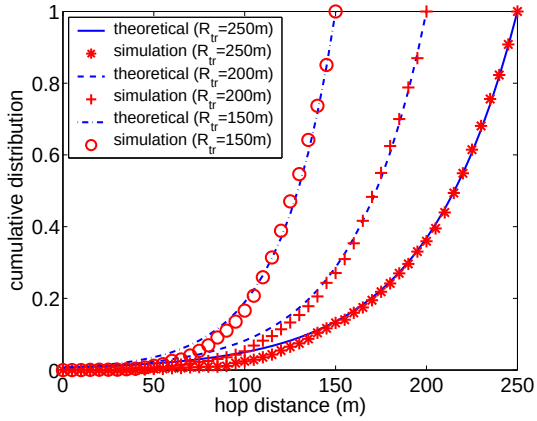


(c) DSDV

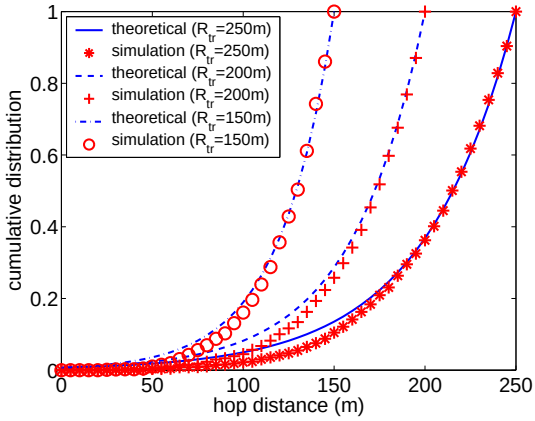
Figure B.3: Average hop distance vs. transmit range in 2D space.



(a) DSDV, n=120



(b) DSDV, n=160



(c) DSDV, n=200

Figure B.4: Hop-distance distribution in 2D space.

Appendix C

The Spatial Reuse Factor of a Chain Traffic Flow

We briefly estimate the spatial reuse factor, η, of a chain traffic flow. In Chapter 3, we use this factor to calculate the throughput upperbound of a chain traffic flow. The method of estimating η is simple. We randomly choose a forwarding node, and estimate the average overlapping area $\overline{o}$ its CSR shares with its closest concurrent transmitters. Let a denote the size of CSR, i.e, $a = 2R_{cs}$ in 1D space. η is

$$\eta = \frac{a}{a - \frac{\overline{o}}{2}} \ . \tag{C.1}$$

For simplicity, forwarding nodes are evenly spaced by the hop-distance d, as shown in Figure C.1. Assume node 0 is transmitting. CSR_0 is of length $2R_{cs}$, and

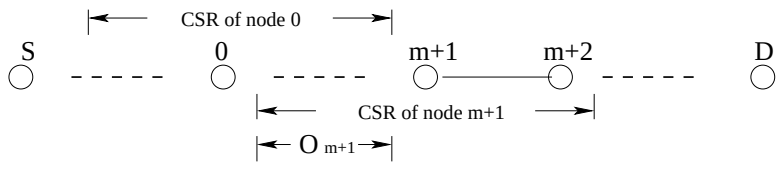


Figure C.1: Spatial reuse factor of a chain traffic flow.

it reserves m forwarding nodes on each side of node 0, i.e.,

$$md \leq R_{cs} < (m+1)d \ . \tag{C.2}$$

On the right side of node 0, because node 1, 2, ..., and m are reserved by CSR_0, they cannot commence simultaneous transmission. Thus, node 0's closest rightside concurrent transmitter is one of node $m+1$, $m+2$, ... $2m+1$. Nodes that are further than $2m+1$ could not overlap CSR_0. Denote o_i as the area CSR_0 overlaps with CSR_i ($i = m+1, m+2, ..., 2m+1$). The distance from the boundary of CSR_0 to node i is $(id - R_{cs})$. Therefore, o_i is

$$o_i = 2R_{cs} - id \ . \tag{C.3}$$

Because node $m+1$, $m+2$, ..., and $2m+1$ compete each other and have to transmit exclusively, each node has $\frac{1}{m+1}$ chances of transmitting concurrently with node 0. Therefore, the average area that CSR_0 shares with its closest rightside concurrent transmitter is

$$\overline{o}_r = \frac{1}{m+1} \sum_{i=m+1}^{2m+1} o_i = 2R_{cs} - \frac{3m+2}{2} d \ . \tag{C.4}$$

So is the area that CSR_0 shares with its closest leftside concurrent transmitter. Thus, the average area that node 0 shares with its closest right and left concurrent transmitters is

$$\overline{o} = 2\overline{o}_r = 4R_{cs} - (3m+2)d \ . \tag{C.5}$$

Substituting $\overline{o}$ into (C.1), yields

$$\eta = \frac{2R_{cs}}{2R_{cs} - (2R_{cs} - \frac{3m+2}{2}d)} = \frac{4R_{cs}}{(3m+2)d} \ . \tag{C.6}$$

According to (C.2), $m \leq \frac{r}{d} < (m+1)$. Therefore, η is bounded by

$$\frac{4m}{3m+2} \leq \eta < \frac{4m+4}{3m+2} \, . \tag{C.7}$$

This shows that, while $m = 2$, $1 \leq \eta < 1.5$. As m increases, η goes to 1.33.

Appendix D

The Spatial Reuse Factor of Parallel Traffic Flows

We use the same method in the previous appendix to estimate η for parallel traffic flows. In Chapter 3, we use this factor to calculate the throughput upperbound of multiple parallel traffic flows. Figure D.1 illustrates multiple parallel traffic flows evenly separated by d_{tf}. d_{tf} is slightly greater than the R_{cs} so that these flows can forward packets concurrently. Assume that the forwarding nodes in each flow is evenly spaced and the hop-distance is d. The number of forwarding nodes reserved by a CSR is $2m + 1$. Thus,

$$m = \left\lfloor \frac{R_{cs}}{d} \right\rfloor \ . \tag{D.1}$$

We randomly chose a traffic flow TF_1. TF_0 and TF_2 are flows above and below TF_1 respectively. Assume that node 10 in TF_1 is transmitting. We consider two kinds of its concurrent transmitters: these within TF_1 and on the rightside of node 10, denoted by node $1i$, and these within TF_2, denoted by $2j$.

First, we derive the average area $\overline{o}_{1r}$ that CSR_{10} overlaps with CSR_{1i}. Node

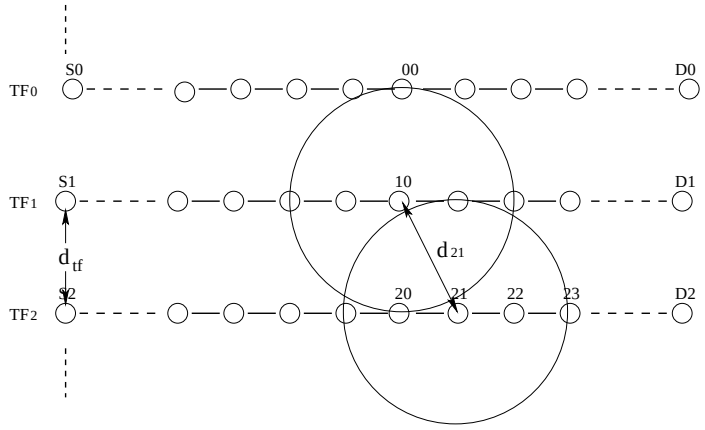


Figure D.1: Spatial reuse factor of multiple parallel traffic flows.

10's closest rightside concurrent transmitter is one of node $m+1$, $m+2$, ... and $2m+1$. The distance from node 10 to node $1i$ is

$$d_{1i} = id \qquad (m < i \leq 2m+1) \ . \tag{D.2}$$

As shown in Figure D.2, the area that CSR_{10} overlaps CSR_{1i} is

$$o_{1i} = R_{cs}^2(\theta_{1i} - sin(\theta_{1i})) \ , \tag{D.3}$$

where

$$\theta_{1i} = \begin{cases} 2arccos(\frac{d_{1i}}{2R_{cs}}) & (R_{cs} < d_{1i} < 2R_{cs}) \\ 0 & (otherwise) \end{cases} \ . \tag{D.4}$$

Note that if $d_{1i} < R_{cs}$, node $1i$ cannot transmit concurrently with node 10; and if $d_{1i} > 2R_{cs}$, CSR_{10} does not overlap with CSR_{1i}. The chances of node $1i$ ($m < i \leq 2m+1$) transmit concurrently with node 10 is $\frac{1}{m+1}$. Therefore, the average area that CSR_0 shares with its rightside concurrent transmitters is

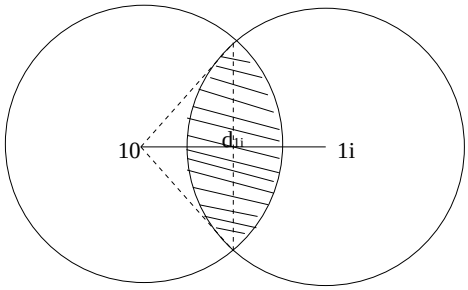


Figure D.2: The overlapping area of CSR_{10} and CSR_{1i}.

$$\begin{aligned}
\overline{o}_{1r} &= \tfrac{1}{m+1}\textstyle\sum_{i=m+1}^{2m+1} o_{1i} \\
&= \tfrac{1}{m+1}\textstyle\sum_{i=m+1}^{2m+1} R_{cs}^2(2arccos(\tfrac{d_{1i}}{2R_{cs}}) - sin(2arccos(\tfrac{d_{1i}}{2R_{cs}}))) \\
&= \tfrac{1}{m+1}\textstyle\sum_{i=m+1}^{2m+1} R_{cs}^2(2arccos(\tfrac{id}{2R_{cs}}) - sin(2arccos(\tfrac{id}{2R_{cs}}))) \\
&= \tfrac{1}{m+1}\textstyle\sum_{i=m+1}^{2m+1} R_{cs}^2(2arccos(\tfrac{i}{2m}) - sin(2arccos(\tfrac{i}{2m})))
\end{aligned} \quad . \tag{D.5}$$

The same result can be obtained for $\overline{o}_{1l}$, which is the average area CSR_{10} shares with its leftside concurrent transmitters.

Second, we derive the average area $\overline{o}_2$ that CSR_{10} overlaps with node $2j$. The euclidian distance from node 10 to node $2j$ is

$$d_{2j} = \sqrt{R_{cs}^2 + (jd)^2} \ . \tag{D.6}$$

Therefore, the overlapping area, o_{2j}, that CSR_{10} shares with CSR_{2j} is

$$o_{2j} = R_{cs}^2(\theta_{2j} - sin\theta_{2j}) \ , \tag{D.7}$$

where

$$\theta_{2j} = \begin{cases} 2arccos(\frac{d_{2j}}{2R_{cs}}) & (R_{cs} \leq d_{2j} \leq 2R_{cs}) \\ 0 & (otherwise) \end{cases} . \quad \text{(D.8)}$$

Note that, if $d_{2j} < R_{cs}$, node $2j$ cannot transmit concurrently with node 10; and if $d_{2j} > 2R_{cs}$, CSR_{2j} does not overlap CSR_{10}. For this reason, we have

$$\begin{aligned} & d_{2j} < 2R_{cs} \\ & \Rightarrow \sqrt{R_{cs}^2 + (jd)^2} < 2R_{cs} \quad . \quad \text{(D.9)} \\ & \Rightarrow |j| \leq \sqrt{3}m \qquad (m \approx \tfrac{R_{cs}}{d}) \end{aligned}$$

Moreover, the chances of node $2j$ transmitting simultaneously with node 10 is $\frac{1}{2m+1}$, because CSR_{2j} covers $2m$ nodes and these nodes' CSR would defer node $2j$'s transmission vice versa. Therefore, the average area that CSR_{10} shares with the concurrent transmitters in TF_2 is

$$\begin{aligned} & \overline{o}_2 = \tfrac{1}{2m+1} \textstyle\sum_{|j| \leq \sqrt{3}m} o_{2j} \\ & = \tfrac{1}{2m+1} \textstyle\sum_{|j| \leq \sqrt{3}m} R_{cs}^2 (2arccos(\tfrac{d_{2j}}{2R_{cs}}) - sin(2arccos(\tfrac{d_{2j}}{2R_{cs}}))) \quad . \quad \text{(D.10)} \\ & = \tfrac{1}{2m+1} \textstyle\sum_{|j| \leq \sqrt{3}m} R_{cs}^2 (2arccos(\sqrt{\tfrac{1}{4} + \tfrac{j^2}{4m^2}}) - sin(2arccos(\sqrt{\tfrac{1}{4} + \tfrac{j^2}{4m^2}}))) \end{aligned}$$

The same result can be obtained for $\overline{o}_0$, which is the average area CSR_{10} shares with the concurrent transmitters in TF_0.

Therefore, the total average overlapping area that CSR_{10} shares with its concurrent transmitters is $\overline{o} = 2(\overline{o}_{1r} + \overline{o}_2)$. Substituting $\overline{o}$ into (C.1) yields

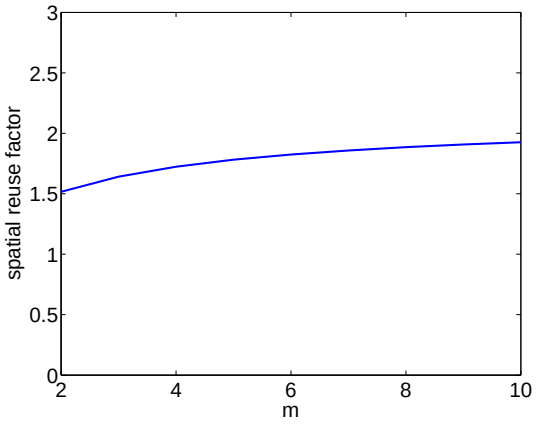


Figure D.3: The spatial reuse factor for the parallel traffic flows.

$$\eta = \frac{\pi R_{cs}^2}{\pi R_{cs}^2 - (\overline{o}_{1r} + \overline{o}_2)} \ . \tag{D.11}$$

Figure D.3 shows that while $m < 4$, η is between 1.5 to 1.7. While m increases, η goes to 1.9.

Bibliography

[1] *IEEE 802.11 Standard for Wireless LAN Medium Access Control and Physical Layer Specifications, (IEEE Computer Society LAN MAN Standards Committee)*, Aug. 1999.

[2] C. E. Perkins, *Ad Hoc Networking*. Addison-Wesley, 2001.

[3] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar, "Next century challenges: Scalable coordination in sensor networks," in *Proc. of the 5th annual ACM/IEEE international conference on Mobile computing and networking (ACM MOBICOM'99)*, Aug. 1999, pp. 263–270.

[4] Wireless mesh network. [Online]. Available: http://en.wikipedia.org/wiki/Wireless_mesh_network/

[5] CMU monarch extensions to NS. [Online]. Available: http://www.monarch.cs.cmu.edu/

[6] NS notes and documentation. [Online]. Available: http://www.isi.edu/nsnam/ns

[7] A. Muqattash and M. Krunz, "POWMAC: a singal-channel power-control protocol for throughput enhancement in wireless ad hoc networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 5, pp. 1067–1084, May 2005.

[8] T. S. Rappaport, *Wireless Communications Principles and Practice*, 2nd ed. Prentice Hall, 2002.

[9] A. Colvin, "CSMA with collision aviodance," Computer Communications, vol. 6, no. 5, pp. 227–35, 1983.

[10] A. Leon-Garcia and I. Widjaja, *Communication Networks: Fundamental Concepts and Key Architectures*, 2nd ed. McGraw-Hill, 2004.

[11] L. L. Peterson and B. S. Davie, *Computer Networks: A Systems Approach*, 3rd ed. Morgan Kaufmann, 2003.

[12] K. Xu, M. Gerla, and S. Bae, "How effective is the IEEE 802.11 RTS/CTS handshake in ad hoc networks?" in *Proc. of Global Telecommunications Conference (IEEE GLOBECOM '02)*, vol. 1, Nov. 2002, pp. 72 – 76.

[13] Y. Zhou and S. M. Nettles, "Balancing the hidden and exposed node problems with power control in CSMA/CA-Based wireless networks," in *Proc. of Wireless Communication and Networking Conference (IEEE WCNC'05)*, Mar. 2005.

[14] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[15] J. G. Proakis, *Digital Communcations*, 4th ed. McGraw-Hill, 2000.

[16] D. R. Pauluzzi and N. C. Beaulieu, "A comparison of SNR estimation techniques for the AWGN channel," *IEEE Transactions on Communications*, vol. 48, no. 10, pp. 1681–1691, Oct. 2000.

[17] M. K. Simon and A. Mileant, "SNR estimation for the baseband assembly," *Telecommunications and Data Acquisition, Report 42-85*, pp. 1681–1691, May 1986.

[18] B. Shah and S. Hinedi, "The split symbol moments SNR estimator in narrow-band channels," *IEEE Transaction on Aerospace and Electronic Systems*, vol. 26, pp. 737–747, Sept. 1990.

[19] H. L. V. Trees, *Detection, Estimation and Modulation Theory.* New York: Wiley, 1968.

[20] R. B. Kerr, "On signal and noise level estimation on a coherent PCM channel," *IEEE Transaction on Aerospace and Electronic Systems*, vol. AES-2, pp. 450–454, July 1966.

[21] R. M. Gagliardi and C. M. Thomas, "PCM data reliability monitoring through estimation of signal-to-noise ratio," *IEEE Transaction on Communications*, vol. COM-16, pp. 479–486, June 1968.

[22] C. M. Thomas, "Maximum likelihood estimation of signal-to-noise ratio," *Ph.D Dissertation*, 1967.

[23] C. E. Gilchriest, "Signal-to-noise monitoring," *JPL Space Programs Summary*, vol. 4, pp. 169–184, June 1966.

[24] T. R. Benedict and T. T. Soong, "The joint estimation of signal and noise from the sum envelope," *IEEE Transactions on Information Theroy*, vol. 13, no. 3, pp. 447–454, July 1967.

[25] R. Matzner, "An SNR estimation algorithm for complex baseband signals using higher order statistics," *Facta Universitatis (Nis)*, no. 6, pp. 41–52, 1993.

[26] R. Matzner and P. Englberger, "An SNR estimation algorithm using fourth-order moments," in *Proc. of 1994 IEEE International Symposium on Information Theory*, June 1994, p. 119.

[27] A. L. Brandão, L. B. Lopes, and D. C. Mclernon, "In-service monitoring of multipath delay and cochannel interference for indoor mobile communication systems," in *Proc. of IEEE International Conference on Communications (IEEE ICC'94)*, vol. 3, May 1994, pp. 1458–1462.

[28] Global positioning system - wikipedia, the free encyclopedia. [Online]. Available: http://en.wikipedia.org/wiki/GPS

[29] S. Capkun, M. Hamdi, and J.-P. Hubaux, "GPS-free positioning in mobile ad-hoc networks," in *Proc. of the 34th Annual Hawaii International Conference on System Sciences*, Jan. 2001, pp. 3481–3490.

[30] R. Iyengar and B. Sikdar, "Scalable and distributed GPS free positioning for sensor networks," in *Proc. of IEEE International Conference on Communications (IEEE ICC'03)*, vol. 1, May 2003, pp. 338–342.

[31] E. M. Royer and C. keong Toh, "A review of current routing protocols for ad hoc mobile wireless networks," *IEEE Personal Communications*, vol. 6, no. 2, pp. 46 – 55, Apr. 1999.

[32] C. Perkins and P. Bhagwat, "Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers," in *Proc. of the Conference on Communications Architectures, Protocols and Applications (ACM SIGCOMM'94)* , vol. 24, 1994, pp. 234–244.

[33] C. Chiang, H. Wu, W. Liu, and M. Gerla, "Routing in clustered multihop, mobile wireless networks with fading channel," in *Proc. of the Singapore International Conference on Networks (IEEE SICON'97)*, Apr. 1997, pp. 197–211.

[34] S. Murthy and J. J. Garcia-Luna-Aceves, "An efficient routing protocol for wireless networks"," *ACM Mobile Networks and Applications*, vol. 1, no. 2, pp. 183–197, Oct. 1996.

[35] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks," in *Mobile Computing*. Kluwer Academic Publishers, 1996, pp. 153–181.

[36] C. E. Perkins and E. M. Royer, "Ad-hoc on-demand distance vector routing," in *Proc. of the 2nd IEEE Workshop on Mobile Computing Systems and Applications (IEEE WMCSA'99)*, Feb. 1999, pp. 90–100.

[37] V. D. Park and M. S. Corson, "A highly adaptive distributed routing algorithm for mobile wireless networks," in *Proc. of the 16th Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM'97)*, vol. 3, Apr. 1997, pp. 1405–1413.

[38] X. Guo, S. Roy, and W. S. Conner, "Spatial reuse in wireless ad-hoc networks," in *Proc. of the 58th Vehicular Technology Conference (IEEE VTC 2003-Fall)*, vol. 3, Oct. 2003, pp. 1437–1442.

[39] L. L. Y. Jing Zhu, Xingang Guo and W. S. Conner, "Leveraging spatial reuse in 802.11 mesh networks with enhanced physical carrier sensing," in *Proc. of International Conference on Communications (IEEE ICC'04)*, vol. 7, June 2004, pp. 4004–4011.

[40] X. Yang and N. H. Vaiday, "On the physical carrier sense in wireless ad hoc networks," in *Proc. of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM'05)*, Mar. 2005.

[41] F. Ye, S. Yi, and B. Sikdar, "Improving spatial reuse of IEEE 802.11 based ad hoc networks," in *Proc. of Global Telecommunications Conference (IEEE GLOBECOM'03)*, vol. 2, Dec. 2003, pp. 1013–1017.

[42] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

[43] N. A. C. Cressie, *Statistics for Spatial Data*. John Wiley & Sons, 1991.

[44] K. L. Yeung and T.-S. P. Yum, "Cell group decoupling analysis of a channel borrowing based dynamic channel assignment strategy in microcellular radio systems," in *Proc. of Global Telecommunications Conference (IEEE Globcom'93)*, vol. 1, Dec. 1993, pp. 281–287.

[45] Y.-B. Lin and V. W. Mak, "Eliminating the boundary effect of a large-scale personal communication service network simulation," *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, vol. 4, no. 2, pp. 165–190, Apr. 1994.

[46] A. G. O. Lugo, F. A. C. Perez, and G. H. Valdez, "Investigating the boundary effect of a multimedia TDMA personal mobile communication network simulation," in *Proc. of the 54th IEEE Vehicular Technology Conference (IEEE VTC-Fall'01)*, vol. 4, Oct. 2001, pp. 2740–2744.

[47] ——, "Impact of the boundary effect and the anticipated channel release strategy in mobile microcellular networks," in *Proc. of the 12th IEEE Internatinal Symposium on Personal, Indoor and Mobile Radio Communications (IEEE PIMRC'01)*, vol. 2, Oct. 2001, pp. F–48 – F–52.

[48] P. K. Jain and H. Haas, "Effects of user distributions on CDMA system performance," in *Proc. of the 16th IEEE Internatinal Symposium on Personal, Indoor and Mobile Radio Communications (IEEE PIMRC'05)*, vol. 3, Sept. 2005, pp. 1652–1656.

[49] C. Bettstetter, "On the minimum node degree and connectivity of a wireless multihop network," in *Proc. of the 3rd ACM International Symposium on Mobile Ad Hoc Networking and Computing (ACM MOBIHOC'02)*, Lausanne, Switzerland, June 2002, pp. 80–91.

[50] C. Bettstetter and J. Zangl, "How to achieve a connected ad hoc network with homogeneous range assignment: an analytical study with consideration of border effects," in *Proc. of the 4th Internatinal Workshop on Mobile and Wireless Communications Network (IEEE MWCN'02)*, Sept. 2002, pp. 125–129.

[51] Z. J. Hass, J. Y. Halpern, and L. Li, "Gossip-based ad hoc routing," in *Proc. of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM'02)*, vol. 3, June 2002, pp. 1707–1716.

[52] S. P. Weber, X. Yang, J. G. Andrews, and G. D. Veciana, "Transmission capacity of wireless ad hoc networks with outage constraints," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 091–4102, Dec. 2005.

[53] X. Yang and G. D. Veciana, "Inducing spatial clustering in MAC contention for spread spectrum ad hoc networks," in *Proc. of the 6th ACM International symposium on Mobile Ad Hoc Networking and Computing (ACM MOBIHOC'05)*, May 2005, pp. 121–132.

[54] A. Hasan and J. G. Andrews, "Scheduling using near-optimal guard zones for CDMA ad hoc networks," in *Proc. of International Conference on Communications (IEEE ICC'06)*, June 2006.

[55] ——, "The critical radius in CDMA ad hoc networks," in *Proc. of Global Telecommunications Conference (IEEE Globcom'04)*, vol. 6, Dec. 2004, pp. 3568–3572.

[56] J. E. Wieselthier and G. D. Nguyen, "Energy-efficient broadcast and multicast trees in wireless networks," *Mobile Networks and Applications*, vol. 7, no. 1, pp. 481–492, Feb. 2002.

[57] J. E. Wieslthier, G. D. Nguyen, and A. Ephremides, "On the construction of energy-efficient broadcast and multicast trees in wireless networks," in *Proc. of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM'00)*, vol. 2, Mar. 2000, pp. 585–594.

[58] T. Chu and I. Nikolaidis, "Energy efficient broadcast in mobile ad hoc networks," in *Proc. of Ad-Hoc Networks and Wireless (ADHOC-NOW))*, Mar. 2002, pp. 84–89.

[59] J. Cartigny, D. Simplot, and I. Stojmenovic, "Localized minimum-energy broadcast in ad-hoc networks," in *Proc. of the 22nd Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM'03)*, vol. 3, Mar. 2003, pp. 2210–2217.

[60] E. L. Lloyd, R. Liu, M. V. Marathe, R. Ramanathan, and S. S. Ravi, "Algorithmic aspects of topology control problems for ad hoc networks," *Mobile Networks and Applications*, no. 1, pp. 19–34, Feb. 2005.

[61] G. Caizzone, P. Giacomazzi, L. Musumeci, and G. Verticale, "A power control algorithm with high channel availability for vehicular ad hoc networks," in *Proc. of the 2005 IEEE International Conference on Communications (IEEE ICC'05)*, vol. 5, May 2005, pp. 16–20.

[62] J.-P. Ebert, B. Stremmel, E. Wiederhold, and A. Wolisz, "An energy-efficient power control approach for WLANs," *Journal of Communications and Networks (JCN)*, vol. 2, no. 3, pp. 197–206, Sept. 2000.

[63] L. M. Feeney and M. Nilsson, "Investigating the energy consumption of a wireless network interface in an ad hoc networking environment," in *Proc. of the 20th Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM'01)*, vol. 3, Apr. 2001, pp. 1548–1557.

[64] C. E. Jones, K. M. Sivalingam, P. Agrawal, and J. C. Chen, "A survey of energy efficient network protocols for wireless networks," *Wireless Networks*, vol. 7, no. 4, pp. 343–358, Aug. 2001.

[65] J. Gomez, A. T. Campbell, M. Naghshineh, and C. Bisdikian, "Conserving transmission power in wireless ad hoc networks," in *Proc. of the 9th International Conference on Network Protocols (IEEE ICNP'01)*, Nov. 2001, pp. 24–34.

[66] B. Chen, K. Jamieson, H. Balakrishnan, and R. Morris, "Span: An energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks," *Wireless Networks*, vol. 8, no. 5, pp. 481–494, Sept. 2002.

[67] E.-S. Jung and N. H. Vaidya, "A power control MAC protocol for ad hoc networks," in *Proc. of the 8th annual International Conference on Mobile Computing and Networking (ACM MOBICOM 2002)*, Sept. 2002, pp. 36–47.

[68] R. Zheng and R. Kravets, "On-demand power management for ad hoc networks," in *Proc. of the 22nd Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM'03)*, vol. 1, Mar. 2003, pp. 481 – 491.

[69] R. Bhatia and M. Kodialam, "On power efficient communication over multi-hop wireless networks: Joint routing, scheduling and power control," in *Proc. of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM'04)*, vol. 2, Mar. 2004, pp. 1457–1466.

[70] U. C. Kozat, I. Koutsopoulos, and L. Tassiulas, "A framework for cross-layer design of energy-efficient communication with QOS provisioning in multi-hop wireless networks," in *Proc. of the 23rd Annual Joint Conference of the IEEE*

Computer and Communications Societies (IEEE INFOCOM'04), vol. 2, Mar. 2004, pp. 1446–1456.

[71] B. Awerbuch, D. Holmer, and H. Rubens, "The pulse protocol: Energy efficient infrastructure access," in *Proc. of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM'04)*, vol. 2, Mar. 2004, pp. 1467–1478.

[72] J. Chen, S.-H. G. Chan, Q. Zhang, W.-W. Zhu, and J. Chen, "A distributed power adaptation algorithm for multimedia delivery over ad hoc networks," in *Proc. of the 2003 International Conference on Multimedia and Expo (ICME'03)*, vol. 1, July 2003, pp. I–293–6.

[73] L. Jia, X. Liu, G. Noubir, and R. Rajaraman, "Transmission power control for ad hoc wireless networks: Throughput, energy and fairness," in *Proc. of Wireless Communications and Networking Conference (IEEE WCNC'05)*, vol. 1, Mar. 2005, pp. 619–625.

[74] V. Kawadia and P. R. Kumar, "Principles and protocols for power control in wireless ad hoc networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 76 – 88, Jan. 2005.

[75] T. ElBatt and A. Ephremides, "Joint scheduling and power control for wireless ad-hoc networks," in *Proc. of the 21rd Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM'02)*, vol. 2, June 2002, pp. 976–984.

[76] A. Behzad, I. Rubin, and A. Mojibi-Yazdi, "Distributed power control medium access control for ad-hoc wireless networks," in *Proc. of 2003 IEEE 18th Annual Workshop on Computer Communications (IEEE CCW'03)*, Oct. 2003, pp. 47–53.

[77] Y. Li and A. Ephremides, "Joint scheduling, power control and routing algorithm for ad-hoc wireless networks," in *Proc. of the 38th Hawaii International Conference on System Sciences (HICSS'05)*, Jan. 2005.

[78] K. Wang, C. F. Chiasserini, R. R. Rao, and J. G. Proakis, "A distributed joint scheduling and power control algorithm for multicasting in wireless ad hoc networks," in *Proc. of International Conference on Communications (IEEE ICC'03)*, vol. 1, May 2003, pp. 725–731.

[79] R. L. Cruz and A. V. Santhanam, "Optimal routing, link scheduling and power control in multihop wireless networks," in *Proc. of the 22th Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM'03)*, vol. 1, Mar. 2003, pp. 702–711.

[80] W. Huang and K. B. Letaief, "Cross-layer scheduling and power control combined with adaptive modulation for wireless ad hoc networks," in *Proc. of Global Telecommunications Conference (IEEE Globcom'05)*, vol. 6, Dec. 2005, p. 5.

[81] S.-L. Wu, Y.-C. Tseng, and J.-P. Sheu, "Intelligent medium access for mobile ad hoc networks with busy tones and power control," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 9, pp. 1647–1657, Sept. 2000.

[82] J. P. Monks, V. Bharghavan, and W. mei W. Hwu, "A power controlled multiple access protocol for wireless packet networks," in *Proc. of the 20th Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM'01)*, vol. 1, Apr. 2001, pp. 219–228.

[83] X. Lin, Y. Kwok, and V. K. N. Lau, "Power control for IEEE 802.11 ad hoc networks: Issues and a new algorithm," in *Proc. of the 2003 International Conference on Parallel Processing (IEEE ICPP'03)*, 2003, pp. 249–256.

[84] F. Alizadeh-Shabdiz and S. Subramaniam, "On the performance of a new 802.11-based low latency power control MAC protocol for ad-hoc networks," in *Proc. of Wireless Communication and Networking Conference (IEEE WCNC'03)*, vol. 2, Mar. 2003, pp. 959–964.

[85] S. Agarwal, S. V. Krishnamurthy, R. H. Katz, and S. K. Dao, "Distributed power control in ad-hoc wireless networks," in *Proc. of the 12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 2, Sept. 2001, pp. F–59 – F–66.

[86] E. Poon and B. Li, "Smartnode: Achieving 802.11 MAC interoperability in power-efficient ad hoc networks with dynamic range adjustments," in *Proc. of the 23th International Conference on Distributed Computing Systems (ICDCS'03)*, May 2003, pp. 650–657.

[87] C. Yu, K. G. Shin, and B. Lee, "Power-stepped protocol: Enhancing spatial utilization in a clustered mobile ad hoc network," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 7, pp. 1322–1334, Sept. 2004.

[88] J. Zhang, Z. Fang, and B. Bensaou, "Adaptive power control for single channel ad hoc networks," in *Proc. of the 2005 IEEE International Conference on Communications (IEEE ICC'05)*, vol. 5, May 2005, pp. 3156–3160.

[89] S. Singh, M. Woo, and C. Raghavendra, "Power-aware routing in mobile ad hoc networks," in *Proc. of the 4th annual ACM/IEEE International Conference on Mobile Computing and Networking (ACM MOBICOM'98)*, Oct. 1998, pp. 181–190.

[90] M. W. Subbarao, "Dynamic power-conscious routing for MANETs: an initial approach," in *Proc. of the 50th Vehicular Technology Conference (IEEE VTC'99-Fall)*, vol. 2, Sept. 1999, pp. 1232–1237.

[91] C. F. Chiasserini and R. R. Rao, "Routing protocols to maximize battery efficiency," in *Proc. of the 21st Century Military Communications Conference Proceedings (IEEE MILCOM'00)*, vol. 1, Oct. 2000, pp. 496–500.

[92] J.-H. Chang and L. Tassiulas, "Energy conserving routing in wireless ad-hoc networks," in *Proc. of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM'00)*, vol. 1, Mar. 2000, pp. 22–31.

[93] Y. Xu, J. Heidemann, and D. Estrin, "Geography-informed energy conservation for ad hoc routing," in *Proc. of the 7th Annual International Conference on Mobile Computing and Networking (ACM MOBICOM'01)*, July 2001, pp. 70–84.

[94] S. Doshi, S. Bhandare, and T. X. Brown, "An on-demand minimum energy routing protocol for a wireless ad hoc network," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 6, no. 3, pp. 50–66, July 2002.

[95] P. Bergamo, A. Giovanardi, A. Travasoni, D. Maniezzo, G. Mazzini, and M. Zorzi, "Distributed power control for energy efficient routing in ad hoc networks," *Wireless Networks*, vol. 10, no. 1, pp. 29–42, Jan. 2004.

[96] J. Zhu, C. Qiao, and X. Wang, "A comprehensive minimum energy routing scheme for wireless ad hoc networks," in *Proc. of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM'04)*, vol. 2, Mar. 2004, pp. 1437–1445.

[97] J.-H. Chang and L. Tassiulas, "Maximum lifetime routing in wireless sensor networks," *IEEE/ACM Transactions on Networking*, vol. 12, no. 4, pp. 609–619, Aug. 2004.

[98] Y.-C. Hu and D. B. Johnson, "Exploiting congestion information in network and higher layer protocols in multihop wireless ad hoc networks," in *Proc. of the 24th International Conference on Distributed Computing Systems (ICDCS'04)*, Mar. 2004, pp. 301–310.

[99] H. Hassanein and A. Zhou, "Routing with load balancing in wireless ad hoc networks," in *Proc. of the 4th ACM international workshop on Modeling, analysis and simulation of wireless and mobile systems*, July 2001, pp. 89–96.

[100] S.-J. Lee and M. Gerla, "Dynamic load-aware routing in ad hoc networks," in *Proc. of IEEE International Conference on Communications (IEEE ICC'01)*, vol. 10, June 2001, pp. 3206–3210.

[101] K. Wu and J. Harms, "Load-sensitive routing for mobile ad hoc networks," in *Proc. of the 10th International Conference on Computer Communications and Networks (IEEE ICCCN'01)*, Oct. 2001, pp. 540–546.

[102] J.-H. Song, V. Wong, and V. C. Leung, "Load-aware on-demand routing (LAOR) protocol for mobile ad hoc networks," in *Proc. of the 57th Vehicular Technology Conference (IEEE VTC'03 Spring)*, vol. 3, Apr. 2003, pp. 1753–1757.

[103] R. Dube, C. D. Rais, K.-Y. Wang, and S. K. Tripathi, "Signal stability based adaptive routing (SSA) for ad-hoc mobile networks," *IEEE Personal Communications*, vol. 4, no. 1, pp. 36 – 45, Feb. 1997.

[104] M. Park, J. G. Andrews, and S. M. Nettles, "Wireless channel-aware ad hoc cross-layer protocol with multiroute path selection diversity," in *Proc. of the 58th IEEE Vehicular Technology Conference (VTC'03-Fall)*, vol. 4, Oct. 2003, pp. 2197–2201.

[105] A. Aduwo and A. Annamalai, "Channel-aware inter-cluster routing protocol for wireless ad-hoc networks exploiting network diversity," in *Proc. of the 60th IEEE Vehicular Technology Conference (VTC'04-Fall)*, vol. 4, Sept. 2004, pp. 2858–2862.

[106] M. R. Souryal, B. R. Vojcic, and R. L. Pickholtz, "Information efficiency of multihop packet radio networks with channel-adaptive routing," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 40–50, Jan. 2005.

[107] D. S. J. D. Couto, D. Aguayo, J. Bicket, and R. Morris, "A high-throughput path metric for multi-hop wireless routing," in *Proc. of the 9th annual international conference on Mobile computing and networking (MOBICOM'03)*, Sept. 2003, pp. 134–146.

[108] R. Draves, J. Padhye, and B. Zill, "Routing in multi-radio, multi-hop wireless mesh networks," in *Proc. of the 10th annual international conference on Mobile computing and networking (MOBICOM'04)*, Sept. 2004, pp. 114–128.

[109] ——, "Comparison of routing metrics for static multi-hop wireless networks," in *Proc. of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications (SIGCOMM'04)*, vol. 34, no. 4, Aug. 2004, pp. 133–144.

[110] C.-K. Toh, "A novel distributed routing protocol to support ad-hoc mobile computing," in *Proc. of the 15th IEEE Annual International Phoenix Conference on Computers and Communications*, Mar. 1996, pp. 480–486.

[111] L. R. Ford and D. R. Fulkerson, *Flows in Networks*. Princeton University Press, 1962.

[112] T. A. ElBatt, S. V. Krishnamurthy, D. Connors, and S. Dao, "Power management for throughput enhancement in wireless ad-hoc networks," in *Proc. of*

IEEE International Conference on Communications (IEEE ICC'00), vol. 3, June 2000, pp. 1506 – 1513.

[113] S. Narayanaswamy, V. Kawadia, R. Sreenivas, and P. Kuma, "Power control in ad-hoc networks: Theory, architecture, algorithm and implementation of the COMPOW protocol," in *Proc. of European Wireless Conference*, Feb. 2002, pp. 156–162.

[114] V. Kawadia and P. Kuma, "Power control and clustering in ad hoc networks," in *Proc. of the 22nd Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM'03)*, vol. 1, Mar. 2003, pp. 459 – 469.

[115] A. Muqattash and M. Krunz, "Power controlled dual channel (PCDC) medium access protocol for wireless ad hoc networks," in *Proc. of the 22nd Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM'03)*, vol. 1, Mar. 2003, pp. 470 – 480.

[116] S. Vural and E. Ekici, "Analysis of hop-distance relationship in spatially random sensor networks," in *Proc. of the 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'05)*, May 2005, pp. 320–331.

[117] Y. chieh Cheng and T. G. Robertazzi, "Critical connectivity phenomena in multihop radio models," *IEEE Transactions on Communications*, vol. 37, no. 7, pp. 770–777, July 1989.

[118] W. C. Lee, "Elements of cellular mobile radio systems," *IEEE Transactions on Vehicular Technology*, vol. VT-35, no. 2, pp. 48–56, May 1986.

[119] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*. The MIT Press, 1990.

[120] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications*, 2nd ed. New York : Wiley, 1995.

[121] L. Kleinrock and F. Tobagi, "Packet switching in radio channels: part 1 - carrier sense multiple-access modes and their throughput-delay characteristics," *IEEE Transactions on Communications*, vol. 23, no. 12, pp. 1400–1416, Dec. 1975.

[122] P. Kern, "MACA - a new channel access method for packet radio," in *Proc. of ARRL/CRRL Amateure Radio 9th Computer Networking Conference*, Sept. 1990.

[123] V. Bharghavan, A. Demers, S. Shenker, and L. Zhang, "MACAW: A media access protocol for wireless LAN's," in *Proc. of the Conference on Communications Architectures, Protocols and Applications (ACM SIGCOMM'94)*, Aug. 1994, pp. 212–225.

[124] C. Fullmer and J. Garcia-Luna-Aceves, "Floor acquisition multiple access (FAMA) in single-channel wireless access," in *Proc. of the Conference on Communications Architectures, Protocols and Applications (ACM SIGCOMM'95)*, vol. 25, no. 4, Oct. 1995, pp. 262–273.

[125] J. Li, C. Blake, D. S. J. D. Couto, H. I. Lee, and R. Morris, "Capacity of ad hoc wireless networks," in *Proc. of the 7th annual International Conference on Mobile Computing and Networking (ACM MOBICOM'01)*, July 2001, pp. 21–29.

[126] A. Shaikh, J. Rexford, and K. G. Shin, "Load-sensitive routing of long-lived IP flows," in *Proc. of the conference on Applications, technologies, architectures, and protocols for computer communication (SIGCOMM'99)*, Aug. 1999, pp. 215–226.

[127] K. C. Claffy, H.-W. Braun, and G. C. Polyzos, "A parameterizable methodology for internet traffic flow profiling," *IEEE Journal on Selected Areas in Communications*, vol. 13, no. 8, pp. 1481–1494, Oct. 1995.

[128] J. Gomez and A. T. Campbell, "A case for variable-range transmission power control in wireless multihop networks," in *Proc. of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM'04)*, vol. 2, Mar. 2004, pp. 1425–1436.

Vita

Yihong Zhou received the degree of Bachelor of Science in Computer Engineering from Beijing University of Posts and Telecommunications (BUPT) in July 1993. Thereafter, she joined Guangzhou Telecommunication Equipment Factory and participated in the development of the first generation China-Made Program-Controlled Telephone Switching Machine. She entered the Graduate School of BUPT in September 1995 and obtained the degree of Master of Science in Computer Engineering. In the following years, she worked for China Telecommunication Corporation, Beijing Branch and participated in several major network construction projects. In June 2000, as the Chief Engineer, she helped establish the VIP Customer Center that provides networking solutions to the most valuable customers. In September 2001, she entered the Graduate School of the University of Texas at Austin.

Permanent Address: 3551 LakeAustin Blvd, Apt D
Austin, TX 78703

This dissertation was typeset with LaTeX 2_ε[1] by the author.

[1] LaTeX 2_ε is an extension of LaTeX. LaTeX is a collection of macros for TeX. TeX is a trademark of the American Mathematical Society. The macros used in formatting this dissertation were written by Dinesh Das and extended by Bert Kay, James A. Bednar, and Ayman El-Khashab.